The physical origin and physiological coding of pinna-based spectral cues

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/32318

Publisher: © Enrique A. López Poveda

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
The physical origin and physiological coding of pinna-based spectral cues

by

Enrique Alejandro López Poveda

A Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy
of Loughborough University
November 1996
© by Enrique A. López Poveda
Esta tesis está dedicada
a mis padres
Abstract

This thesis investigates both the physical origin and the physiological coding of pinna-based spectral features observed in experimental Head-Related Transfer Functions (HRTFs). An experimental method for measuring HRTFs using a Knowles Electronics Manikin for Acoustic Research (KEMAR) is presented. The method includes a technique for moulding individualised pinnae to be fitted on to the KEMAR's head. Experimental HRTF data obtained with this method are shown and analysed. The most remarkable pinna-based spectral features are identified and their dependency on source location is characterised, particularly for elevation-dependent spectral notches. A physical model of sound diffraction, multiple reflections and interference in the human concha is developed and tested. The physical processes in the concha underlying the generation of characteristic elevation-dependent spectral notches are investigated with the proposed model. The model indicates that sound diffraction in the concha cavity explains the fact that the same elevation-dependent spectral notches are observed for all vertical planes, with very little azimuthal variation.

A computer model of the activity of low-, medium- and high-spontaneous rate auditory nerve fibres is also developed and tested. The model includes the following stages: stimulus generation, middle-ear filtering, non-linear auditory filter bank, inner hair-cell (IHC) receptor potential, transmitter-release function, inner hair cell-auditory nerve synapse, and spike generation and refractory effects. Special effort has been dedicated to obtain a good description of the relationship between the IHC receptor potential and the membrane permeability. The model simulates experimental rate-intensity curves of auditory nerve fibres both for stimuli at- and off- the fibre's best frequency for best frequencies within the range 4 to 18 kHz. The physiological coding of pinna-based spectral features is investigated by using the proposed model to simulate auditory nerve rate-profiles in response to HRTF-filtered broad-band stationary stimuli. The model shows that the quality of the auditory-nerve representation of the pinna-based spectral features deteriorates with signal level. Moreover, the model indicates the broadening of the cochlear filters as a function of stimulus intensity is the factor mostly responsible for the deterioration of the AN rate profiles at high signal levels. Transient rate-profiles for moving sources are also investigated with the model. The model
indicates that for signal levels within the dynamic range of the fibre, the transient response between two locations (i.e., between two HRTFs) resembles the first order differential of the two corresponding HRTFs. Possible mechanisms for coding the spectral information present in HRTFs are proposed and discussed.
I wish to express my most sincere thanks to Prof. Ray Meddis, for without his constant advice and support I would not know anything about hearing or research. I could not have had a better supervisor.

I would also like to express my gratitude to all members of the Speech and Hearing Lab. at Loughborough University; in particular, to Stuart Hunter for his technical support; to Lowel O'Mard for his patience with my unorthodox programming style and his help with the computing aspects of the work; and to Emily Shotter for her cheerful personality and pleasant company.

Additionally, I would like to thank all the staff in the Department of Human Sciences at Loughborough University. Thank you for giving me the chance to learn with you and for all your help and advice throughout the last three years.

I owe a great deal of thanks to all my friends in Loughborough and Spain, who, in their own way, kept me going. You know who you are, and I appreciate your friendship and hope it continues.

This work would not have been possible without the constant support of my whole family but especially of three very special people in my life: my parents and Almu. Thank you, Almu, for all your help and support and for keeping me company during these difficult years. Thank you, mum and dad, for without your financial and emotional support this thesis would not have been possible. It is to you that I dedicate this thesis.
Table of contents

Abstract ................................................................................................................................... iv
Acknowledgements ........................................................................................................ vi
Table of contents ........................................................................................................ vii
List of abbreviations and acronyms .............................................................................. xv
List of symbols .............................................................................................................. xvii
List of figures ............................................................................................................... xviii
List of tables ................................................................................................................. xxix

CHAPTER 1
Introduction ...................................................................................................................... 1
  1.1 Motivation .................................................................................................................. 1
  1.2 Objectives ................................................................................................................ 2
  1.3 Original contributions ............................................................................................. 5
  1.4 Overview of the thesis ............................................................................................ 7
  1.5 Angle conventions ................................................................................................... 11

CHAPTER 2
Background ...................................................................................................................... 12
  2.1 Introduction .............................................................................................................. 12
    Explaining HRTFs by studying the transfer function of individual anatomical features ..... 13
  2.2 Source-location information encoded in binaural HRTFs ....................................... 14
    2.2.1 Binaural information ....................................................................................... 15
      The effect of the head: Interaural time and level differences .................................. 15
      A simple geometrical model of ITDs ................................................................ 15
      The pressure field at any point on a rigid sphere ................................................. 20
      Limitations of the "Duplex" theory of sound localisation ................................... 23
    Pinna disparity: another source of binaural information in the form of ILDs ........ 24
2.2.2 Monaural information: pinna-based spectral cues .......... 25
   The nature of the location-dependent information
   provided by the filtering action of the pinna: temporal or
   spectral cues? ................................................. 25
   Identification of pinna-based spectral cues in HRTFs .... 28
   The role of pinna-based spectral cues in sound
   localisation ................................................... 29
   The problem of developing a mathematical description
   of the transfer function of the pinna ................................................... 29

2.3 Auditory nerve encoding of pinna-based spectral cues .......... 31
   2.3.1 Rate representations of pinna-based spectral features .... 32
   Experimental results and considerations .................... 32
   Stationary stimuli ...................................... 32
   Moving stimuli .......................................... 33
   2.3.2 Modelling the AN rate response to HRTF-filtered stimuli . 34

CHAPTER 3
Experimental head-related transfer functions measured using a
KEMAR fitted with individualised moulded pinnae .............. 36
   3.1 Introduction ................................................................. 36
   3.2 Experimental technique for moulding real pinnae ............ 37
      3.2.1 Health and safety measures .................................... 38
      3.2.2 Making the 'negative' pinna mould ......................... 38
      3.2.3 Making the pinna impression .................................. 38
      3.2.4 Fixing the pinna impression to the KEMAR mould ...... 39
      3.2.5 Pinna replacements: the 'infills' ......................... 40
   3.3 Experimental method and equipment for measuring HRTFs .... 40
      3.3.1 Stimulus generation: the super-pulse .................... 42
      Other measures adopted to improve the quality of the
      super-pulse ...................................................... 45
      3.3.2 HRTF measurements ........................................... 46
   3.4 System validation: Results ........................................ 47
      3.4.1 HRTF measurements with the pinnae in place .......... 48
      Results: HRTF measurements with the pinna in place .... 48
3.4.2 The transfer function of the KEMAR head and torso:
HRTFs measurements without the pinnae ......................... 53
Results: HRTFs measurements without the pinnae ........ 53
3.4.3 An approximated experimental pinna transfer-function.... 56
3.5 Discussion.................................................................. 59
3.5.1 Experimental method ........................................... 59
3.5.2 Spectral features................................................... 60

CHAPTER 4
Experimental head-related transfer functions measured using a
KEMAR fitted with standard pinnae ........................... 62
4.1 Introduction ................................................................. 62
4.2 Analysis method for obtaining the HRTFs ...................... 63
4.3 Results....................................................................... 64
4.3.1 A comparison between the HRTFs obtained with
individualised moulded pinnae and those obtained with
standard DB-061 pinnae ............................................ 64
4.3.2 Elevation dependency ........................................... 65
4.3.3 Azimuthal dependency ........................................... 72
4.4 Discussion and Conclusions ........................................... 75
4.4.1 Azimuthal dependency .......................................... 75
4.4.2 Elevation dependency ........................................... 75
4.4.3 Generalisation of the results to other pinnae .......... 76
4.4.4 Generalisation of the results to human HRTFs......... 78
4.4.5 Psychophysical implications ................................... 78
4.4.6 The physical cause of the most salient spectral notches .... 79

CHAPTER 5
A physical model of sound diffraction and reflections in the
human concha............................................................... 83
5.1 Introduction ................................................................. 83
5.2 The diffraction/reflection model: Theory ......................... 85
5.2.1 General ........................................................... 86
5.2.2 Theoretical approximations and assumptions ............ 87
5.2.3 Diffracted pressure at any point on the concha reflecting
wall ........................................................................... 90
5.2.4 Reflected pressure at the meatus entrance .................... 91
5.2.5 The transfer function of the concha ............................... 92
5.3 Validation of the theory .................................................... 93
  5.3.1 Experimental method ........................................... 93
  5.3.2 Results ............................................................ 94
  5.3.3 A comparison between the performance of the single-
delay-and-add approximation and the diffraction/reflection
  model ..................................................................... 97
  Two-dimensional approximation ........................................ 99
5.4 Predictions of the diffraction/reflection model for a realistic
concha shape .................................................................................. 99
5.5 Discussion ........................................................................ 105
5.6 Conclusions ....................................................................... 107

CHAPTER 6
The contribution of individual anatomical features of the pinna to
experimental head-related transfer functions ............................. 109
  6.1 Introduction ................................................................... 109
  6.2 Experimental method ..................................................... 111
    6.2.1 Experimental set-up and equipment: An improved super-
pulse method................................................................. 111
    6.2.2 Procedure .................................................................. 115
  6.3 Results ......................................................................... 119
    6.3.1 Unaltered pinna .................................................. 119
    6.3.2 Flange-filled condition .......................................... 121
    6.3.3 No-flange condition ............................................ 123
    6.3.4 Concha without the crus helias ................................ 124
    6.3.5 Concha without the crus helias and without the tragus ..... 124
  6.4 Discussion ..................................................................... 129
    6.4.1 Compatibility of the results with the predictions of the
diffraction/reflection model ................................................ 130
    6.4.2 Compatibility of the results with related research
previously published .......................................................... 131
    6.4.3 Psychophysical implications ................................... 132
  6.5 Summary ....................................................................... 133
CHAPTER 7
A computer model of the auditory nerve response to HRTF stimuli using a linear gamma-tone auditory filter bank .......... 134

7.1 Introduction ................................................................. 134

7.2 Model description ........................................................... 136
7.2.1 Stages (1), (2) and (3): HRTF-stimulus generation ...... 138
7.2.2 Stage (4): Gamma-tone filter bank ......................... 139
7.2.3 Stage (5): The mechanical to neural transduction .......... 140
7.2.4 Stage (6): Spike generation and refractory effects ....... 143

7.3 Parameters for the IHC model: Rate-intensity curves ......... 144

7.4 Model evaluation: Auditory-nerve rate profiles in response to HRTF stimuli ............................................................. 149
7.4.1 Output profiles from the gamma-tone filter bank ........... 149
7.4.2 Onset and steady-state rate profiles for HSR fibres in response to HRTF-filtered broad-band noise ...................... 152
7.4.3 Rate profiles for LSR, MSR and HSR fibres at different stimulus intensities ...................................................... 156

7.5 Discussion and Conclusions .............................................. 160
7.5.1 Model limitations ....................................................... 160
7.5.2 Diversity of AN fibres: Implications in modelling the AN response to HRTF stimuli ........................................ 161
7.5.3 Psychophysical implications for sound localisation ........ 162

CHAPTER 8
A computer model of low, medium and high spontaneous rate auditory-nerve fibres using a dual-resonance nonlinear auditory filter bank ........................................................... 163

8.1 Introduction ................................................................. 163

8.2 Model description ........................................................... 166
8.2.1 Stage (1): Stimulus generation ................................. 168
8.2.2 Stage (2): Middle-ear filtering ..................................... 169
8.2.3 Stage (3): Dual-resonance nonlinear (DRNL) auditory filter bank ................................................................. 171

DRNL model overview ..................................................... 171
Creating a filter-bank model .............................................. 174
Nonlinear phenomena simulated by the DRNL model ......... 175
8.2.4 Stage (4): BM displacement to IHC stereocilia displacement coupling ........................................ 179
8.2.5 Stage (5): IHC stereocilia displacement to IHC receptor potential coupling ........................................ 180
   Model overview ........................................ 180
   Transfer characteristics of the model ................. 184
   Model parameters ........................................ 184
8.2.6 Stage (6): Transmitter release ......................... 185
8.2.7 Stage (7): Synaptic effects ........................................ 186
8.2.8 Stage (8): Spike generation and refractory effects........ 187
8.3 Model evaluation and parameters selection ............ 187
   8.3.1 IHC intracellular receptor potential .................. 188
   DC receptor potential as a function of stimulus intensity .... 188
   AC/DC ratio as a function of stimulus frequency .......... 192
8.3.2 AN response: LSR, MSR and HSR rate-intensity curves for stimuli at BF ........................................ 196
   Model evaluation procedure and results ............. 196
   Model parameters: Differences between LSR, MSR and HSR fibres ........................................ 197
   Simulated RI curves across BFs ...................... 200
8.3.3 AN response: Rate-intensity curves for stimuli off BF .... 205
   Model parameters ................................... 205
   Results: off-BF RI curves .............................. 206
8.4 Discussion and Conclusions .......................... 210
   8.4.1 Model parameters ........................................ 210
   8.4.2 Low, Medium and High spontaneous rate AN fibres.... 211
   8.4.3 Model applications ...................................... 214

CHAPTER 9
Simulating auditory nerve responses to HRTF stimuli using a nonlinear auditory filter bank .......................... 215
   9.1 Introduction ........................................ 215
   9.2 Stimulus generation ........................................ 217
   9.3 Model parameters, Simulated rate-intensity curves and Evaluation protocol ........................................ 220
   9.3.1 Rate-intensity curves for broad-band noise .......... 221
9.3.2 Model evaluation protocol ................................................. 222
9.4 Results: Stationary stimuli ................................................ 225
  9.4.1 Output profiles from the DRNL filter bank .................. 225
  9.4.2 Onset and steady-state rate profiles associated with HSR
fibres ........................................................................ 228
  9.4.3 Rate profiles of LSR, MSR and HSR fibres at different
stimulus intensities .................................................. 229
9.5 Results: Moving stimulus ................................................. 233
9.6 Discussion ................................................................... 239
  9.6.1 Cochlear output profiles ........................................ 239
  9.6.2 AN rate profiles for stationary stimuli ................ 241
  9.6.3 Moving Stimuli ................................................. 242
9.7 Summary and Conclusions ................................................ 243

CHAPTER 10
General discussion and conclusions ....................................... 245
  10.1 Summary ................................................................... 245
  10.2 Discussion and Conclusions ............................................. 246
    10.2.1 Technical considerations: An experimental method for
making HRTF measurements using a KEMAR fitted with
individualised moulded pinnae ........................................ 246
      Pinna-moulding technique ........................................ 246
      An experimental method for measuring HRTFs ............ 247
    10.2.2 The location dependency of pinna-based spectral
features ........................................................................ 247
    10.2.3 Modelling the transfer function of the pinna: The
physical origin of pinna-based spectral features ............... 248
    10.2.4 A computer model of the auditory peripheral system.... 250
    10.2.5 The physiological coding of pinna-based spectral cues .. 252
      Stationary stimuli .................................................. 252
      Moving stimuli .................................................... 253
    10.2.6 The ambiguity of spectral cues in sound localisation:
What happens if the sound source does not have a flat
spectrum? .............................................................. 254
  10.3 Proposed future work .................................................... 255

xiii
List of abbreviations and acronyms

Some abbreviations and acronyms are used in this thesis. Here is a list of them presented in alphabetical order:

AC: Alternating Current.
ADC: Analogue-to-Digital Converter.
AIFF: Audio Interchange File Format.
AN: Auditory Nerve.
BF: Best Frequency.
CF: Centre Frequency.
cm: centimetre.
CNS: Central Nervous System.
DAC: Digital-to-Analogue Converter.
dB SPL: dB (re 20x10⁻⁶ Pa).
dB: decibel.
DC: Direct Current.
DRNL: Dual-Resonance Non-Linear (auditory filter model).
Eq: Equation.
FFT: Fast Fourier Transform.
Fig: Figure.
g: grams.
HP: Horizontal Plane.
HRIR: Head-Related Impulse Response.
HRTF: Head-Related Transfer Function.
HSR: High Spontaneous Rate.
Hz: Hertz (cycles per second).
I/O: Input/Output (function).
IFFT: Inverse Fast Fourier Transform.
IHC: Inner Hair Cell.
ILD: Interaural Level Difference.
IPD: Interaural Phase Difference.
ITD: Interaural Time Difference.
KEMAR: Knowles Electronics Manikin for Acoustic Research.

kHz: kilohertz.

LSR: Low Spontaneous Rate.

LVP: Lateral Vertical Plane.

m: metres.

min: minutes.

ml: millilitres = 10^-3 l.

ms: millisecond = 10^-3 s.

MSR: Medium Spontaneous Rate.

mV: millivolt = 10^-3 V.

MVP: Median Vertical Plane.

Pa: Pascals (pressure unit equivalent to N/m^2).

PSTH: Post-Stimulus Time Histogram.

RI: Rate-Intensity (refers to Rate-Intensity curves).

rms: root mean square.

s: second.

SPL: Sound Pressure Level.

SR: Spontaneous Rate.

μPa: micro-Pascal = 10^-6 Pa.

μs: microsecond = 10^-6 s.
List of Symbols

This list only contains variables that have not been defined in the main text. Complex magnitudes have been written in bold characters, e.g., $p_0$. A small bar on the character indicates vectorial notation, e.g., $\bar{k}$.

General symbols

- $\phi$: angle of elevation.
- $\theta$: azimuthal angle.
- $f$: frequency.
- $t$: time.
- $\omega$: angular frequency; $\omega = 2\pi f$
- $k$: wave vector; $k = \frac{2\pi}{\lambda}$
- $\lambda$: wavelength.
- $j$: $= \sqrt{-1}$.
- $\rho$: density of the propagating medium.
- $c$: speed of sound. We have considered $c = 332.1$ m/s.

Chapter 5 and Appendix 1

- $P_T$: total pressure at the meatus entrance.
- $P_R$: reflected pressure from the concha wall measured at the meatus entrance.
- $P_U$: $= P_T - P_R$
- $p_0(\bar{r},t)$: incident sound pressure at point $\bar{r}$.
- $p_0$: amplitude of the incident sound wave.
- $\bar{n}_l$: unit vector normal to the diffracting aperture.
- $\bar{n}_q$: unit vector normal to the reflecting surface.
- $\bar{u}(\bar{r},t)$: particle velocity at point $\bar{r}$. It is defined as: $\bar{u}(\bar{r},t) = -\frac{1}{\rho} \int \nabla p(\bar{r},t) dt$

(Morse and Ingard, 1968).
CHAPTER 2

Figure 2.1. (a) 3D Geometrical model of ITDs. The head is approximated by a perfect sphere and the ears are regarded as points on both sides of the head (L and R for the left and right ears respectively). The sound source is indicated by $P(x_p,y_p,z_p)$ in Cartesian co-ordinates or $P(d,\theta,\phi)$ is spherical co-ordinates. (b) 2D geometrical model of ITDs. Notice that this representation is equivalent to (a) but with $\phi=0$. In both models, the interaural time difference is calculated as $\text{ITD} = (D_L - D_R)/c$.

Figure 2.2. ITDs for a spherical head (of radius $a = 9.3$ cm) as predicted by a simple geometrical model (see text for details). (a) ITD as a function of the azimuthal angle, for 10 different elevations between $0^\circ$ and $90^\circ$ in steps of $10^\circ$. (b) ITD as a function of the elevation angle, for 10 different azimuths between $0^\circ$ and $90^\circ$ in steps of $10^\circ$. The results presented in (a) and (b) are for a source $1$ m away from the centre of the sphere. (c) ITD as a function of source distance, for azimuths between $0^\circ$ to $90^\circ$ in steps of $15^\circ$. Distances are measured from the centre of the sphere.

Figure 2.3. Intensity gain (in dB) at the left-ear point in a spherical head. The values were calculated as $20\log_{10}\left(\frac{(p_1+p_2)}{p_0}\right)$, where $(p_1+p_2)/p_0$ was obtained from evaluating Eq. (2.11), (see text for details). $\theta$ represents the azimuthal angle and $\alpha$ represents the angle between the incident direction of the sound and the point at which the pressure gain is being measured (see inset on bottom-right corner).

Figure 2.4. The external ear (adapted from Shaw, 1975b). Notice the three major sub-divisions of the external ear: the meatus (dark pattern), the concha (light pattern) and the pinna-flange (non-shaded area).
CHAPTER 3

Figure 3.1. Experimental set-up and equipment for HRTF measurements.

Figure 3.2. Super-pulse generation technique. (a) Time-domain average response of system to a specified number of Dirac pulses. (b) Spectrum of the average time response shown in (a). (c) Spectral mirror image of (b). (d) The super-pulse: time-domain signal corresponding to the mirror spectrum shown in (c). (e) Average system's response to N super-pulses. (f) Spectrum of the average system's response shown in (e). Note that the frequency range is from 1 to 10 kHz in this case. Note the spectrum of the average response of the system to N super-pulses (f) has spectral variations of ±1 dB in this case.

Figure 3.3. HRTFs measured with the KEMAR fitted with the pinnae of subject IX as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Notice that the total pressure at the microphone (eardrum simulator) is plotted instead of the actual transfer function (see text for details). Notice that the main spectral features are marked: P = peak, N = notch. N.B., the frequency range goes from 1 to 10 kHz.

Figure 3.4. HRTFs measured with the KEMAR fitted with the pinnae of subject AC as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Legend as in Fig. 3.3.

Figure 3.5. HRTFs measured with the KEMAR fitted with the pinnae of subject LPO as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Legend as in Fig. 3.3. N.B. the frequency range goes from 1 to 17 kHz.

Figure 3.6. HRTF for the KEMAR fitted with the infills (i.e., no pinnae) as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Notice the total pressure at the microphone (eardrum simulator) is plotted instead of the actual transfer function (see text for details). Notice that the main spectral features are marked: P = peak, N = notch. N.B. the frequency range is from 1 to 17 kHz.
Figure 3.7. Approximated transfer function of the pinna calculated by subtracting
the infill HRTF (Fig. 3.6) from the HRTF data for subject LPO (Fig. 3.5). (a) Left
pinna. (b) Right pinna.

CHAPTER 4

Figure 4.1. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the azimuthal angle, for a source at $-40^\circ$ elevation. The HRTF
data were obtained by analysing the HRIR database provided by Gardner and Martin

Figure 4.2. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the elevation angle, for a source in the median vertical plane ($0^\circ$
azimuth). The HRTF data were obtained by analysing the HRIR database provided by

Figure 4.3. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the elevation angle, for a source in the lateral vertical plane ($270^\circ$
azimuth). The HRTF data were obtained by analysing the HRIR database provided by

Figure 4.4. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the elevation angle, for a source in the ipsilateral diagonal vertical
plane ($315^\circ$ azimuth). The HRTF data were obtained by analysing the HRIR database

Figure 4.5. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the elevation angle, for a source in the contralateral diagonal
vertical plane ($45^\circ$ azimuth). The HRTF data were obtained by analysing the HRIR

Figure 4.6. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-
061) as a function of the azimuthal angle, for a source in the horizontal plane ($0^\circ$
Figure 4.7. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimuthal angle, for a source at +40° elevation. The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.

Figure 4.8. HRTFs measured at the right eardrum of a KEMAR (pinna model DB-065) as a function of the elevation angle, for a source in the median vertical plane (0° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.

Figure 4.9. HRTFs measured at the right eardrum of a KEMAR (pinna model DB-065) as a function of the elevation angle, for a source in the ipsilateral diagonal vertical plane (45° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.

Figure 4.10. HRTFs measured at the right eardrum of a KEMAR (pinna model DB-065) as a function of the elevation angle, for a source in the lateral vertical plane (90° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.

CHAPTER 5

Figure 5.1. (a) A two-dimensional view of the geometry of the problem and convention taken for the definition of the different variables involved. (b) Experimental set-up and dimensions of the reflecting spiral-shaped surface.

Figure 5.2 A comparison of the experimental data with the predictions of the single-delay-and-add approximation and the proposed diffraction/reflection model, at three elevations (−45°, −17° and +35°), for a metal spiral-shaped reflecting surface whose dimensions are shown in Fig. 5.1b. The thin continuous line shows the experimental results. The heavy line shows the predictions of the diffraction/reflection
model when no attenuation with distance of the diffracted wave is considered. The heavy dashed line in the third panel (-45° elevation) shows the predictions of the diffraction/reflection model when attenuation with distance of the diffracted wave is considered. The arrows indicate the cancellation frequencies as predicted by the single-delay-and-add approximation for the metallic spiral.

**Figure 5.3.** (a) The inner contour (lateral view or YZ projection) of a real concha which was used to generate the approximate 3D-concha shape for which the diffraction/reflection theory has been evaluated. (b) Top view (or XY projection) of the approximate three-dimensional concha. Note that the protrusion angle is 158° and that the concha cavity is effectively described by two reflecting surfaces ($S_L$ and $S_P$) and by the diffracting aperture $S_A$. The axes units are given in cm. and express to scale the real dimensions of the concha.

**Figure 5.4.** Top: HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle of a sound source placed in the median vertical plane. This figure is a reproduction of Fig. 4.2. Notice that the data include the effect of all the body parts which modify the incident sound before it reaches the eardrum; i.e., torso, head, and the whole of the external ear. Bottom: Predictions of the diffraction/reflection model for a realistic concha shape (see main text for details). Only the spectral features associated with the tranverse dimensions of the concha are plotted in this panel.

**Figure 5.5** Top: HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimuth angle for a sound source placed at -40° elevation. This figure is a reproduction of Fig. 4.2. Bottom: The predictions of the diffraction/reflection model for a realistic concha shape (see main text for details).

**CHAPTER 6**

**Figure 6.1.** Spectra of the average response of the system to 40 Dirac pulses (thin line) and 40 super-pulses (thick line) —see text for details—. Note that the response of the system to the super-pulse is almost flat with maximum variations of ±1.5 dB between 2 and 17 kHz.
Figure 6.2. (a) Unaltered LPO left pinna (see also Appendix 1). (b) Flange-filled ear-piece. (c) No-flange ear-piece. Notice that the flange has been removed. (d) Concha-no-crus-helia ear-piece. Notice that the crus helias (see Fig. 2.4) has been removed from the 'no-flange' ear-piece shown in (c). (e) Concha-no-tragus ear-piece. Notice the tragus has been removed from the ear-piece shown in (d).

Figure 6.3. HRTF results after altering the anatomical structure of the pinna as a function of source azimuth. The elevation angle was $-45^\circ$. (a) unaltered pinna. (b) flange-filled. (c) no-flange. (e) concha-no-crus-helia. (d) concha-no-tragus. The most important differences between successive conditions are marked with an asterisk. Corresponding spectral features are similarly labelled. (see text for details).

Figure 6.4. As in Fig. 6.3 but for a source at $0^\circ$ elevation.

Figure 6.5. As in Fig. 6.3 but for a source at $+45^\circ$ elevation.

CHAPTER 7

Figure 7.1. Processing sequence of the model.

Figure 7.2. Diagram illustrating the operation of the (synaptic effects) hair cell model [Adapted from Meddis et al. (1990)].

Figure 7.3. (a) Onset rate-intensity curve simulated by the model using the set of parameters given in Meddis et al. (1990) for a HSR fibre — see Table 7.1 —. (b) Steady-state rate-intensity curve simulated by the model with the same set of parameters. N.B. different scales for y-axis.

Figure 7.4. (a) Simulated steady-state rate-intensity curves. (b) Simulated onset rate-intensity curves. (c) Experimental rate-intensity curves [reproduced from Winter and Palmer (1991)]. CF = 2.1 kHz. The modelled curves were generated with the new sets of parameters for the mechanical-to-neural transduction model (Meddis, 1986, 1988) that are given in Table 7.1. Notice that the spontaneous rate of the
experimental and modelled curves is approximately the same. However, the saturation rate is different (see text for details).

**Figure 7.5.** Output from the linear gamma-tone filter bank when the stimulus is a Dirac pulse (click) convolved with the HRIR corresponding to a sound source at \(-20^\circ\) elevation, 0° azimuth. Three different stimulus intensities are represented 30, 50, and 80 dB SPL (rms). The bottom series represents the spectrum of the HRIR. N.B. the x-axis units must be understood as frequency (kHz) for the HRIR spectrum but centre frequency (kHz) for the gamma-tone output profiles.

**Figure 7.6.** (a) Spectrum of the HRIR corresponding to a sound source at 0° azimuth, and \(-20^\circ\) elevation. (b) Onset rate profile at four signal levels: 35, 50, 85 and 110 dB SPL (rms). The onset rate for each CF represents the maximum rate over the first 10 ms of the stimulus duration (50 ms). (c) Steady-state rate profiles for the same stimulus and intensities as in b). The steady-state rate for each CF represents the average over the last 20 ms of the 50 ms noise burst. (d) Average rate over the full duration of the noise burst (50 ms). N.B. different y-axis scales.

**Figure 7.7** Approximate rate profiles for the three fibre types, LSR, MSR and HSR, in response to a click with the spectrum of the HRIR shown in Fig. 7.6a. The maximum discharge rate is shown in the y-axis (see main text for details). (a) 30 dB SPL signal level. (b) 50 dB SPL signal level. (c) 80 dB SPL signal level.

**CHAPTER 8**

**Figure 8.1.** Processing sequence of the model.

**Figure 8.2.** (a) The shape of the pre-emphasis filter (\(\alpha_{PE} = 0\) dB). (b) Model middle-ear threshold tuning curve compared with the psychophysical audiograms for the guinea pig (Prosen et al., 1978) and chinchilla (Clark and Bohne, 1986).

**Figure 8.3.** Schematic of the DRNL filter model. Adapted from figure 1 of O'Mard and Meddis (1996).
Figure 8.4. Two-tone suppression phenomena at 8 kHz BF (adapted from figure 6 of O'Mard and Meddis, 1996). (a) The DRNL model response. (b) Re-plotted from auditory-nerve data by Arthur et al. (1971). The dark-shading represents suppression areas, whereas the light-shading represents the excitation area.

Figure 8.5. Two-tone combination tones (adapted from figure 7 of O'Mard and Meddis, 1996). BF = 34 kHz, \( f_1 = 36 \) kHz, and \( f_2 = 38 \) kHz. (a) DRNL model response. (b) Experimental data for the cat basilar membrane (figure 5a of Cooper and Rhode, 1995).

Figure 8.6. Schematic diagram of a radial section of the cochlea, including the equivalent circuit of the inner hair cell (shaded box) and surrounding structures. Adapted from figure 4 of Shamma et al. (1986). \( E_t = \) endocochlear potential; \( R_e, R_p = \) epithelium resistances; \( E_K = \) potassium reversal potential; \( G_K = \) ionic channel conductance of the hair cell basal membrane; \( G_m(u) = \) mechanically sensitive conductance of the hair cell apical membrane; \( G_a = \) leakage conductance of the hair cell apical membrane; \( u = \) cilia displacement; \( V = \) IHC intracellular potential.

Figure 8.7. (a) Experimental and simulated DC component of the receptor potential as a function of stimulus intensity for two stimulus frequencies (18 and 7 kHz). The BF of the DRNL filter is 18 kHz. The symbols represent two sets of experimental data (adapted from figure 6a of Patuzzi and Sellick, 1983). The lines represent the model results obtained with the parameters given in Table 8.2 (i.e., \( \alpha p_E = -29 \) dB SPL and \( \Theta_{Th} = 15.47 \) dB SPL). The stimulus frequency is indicated by the numbers by each line. (b) DRNL filter output for a pure-tone stimulus at BF (continuous line), and off-BF (long-dashed line), with the set of parameters given in Table 8.2. The BF of the filter was 18 kHz. The numbers by each line indicate the stimulus frequency. The thin dotted-line illustrates a linear relationship. (c) and (d) Same as (a) and (b) but with \( \alpha p_E = 0 \) dB SPL and \( \Theta_{Th} = 29.5 \) dB SPL; i.e., with the DRNL parameters originally proposed by O'Mard and Meddis (1996).

Figure 8.8. A comparison between experimental and simulated AC/DC ratios as a function of stimulus frequency. The experimental curves (black symbols) have been adapted from figure 10 of Palmer and Russell (1986). The model response is given by the white squares linked by a thick line. The BF of the cell was 18 kHz. The
model parameters are given in Table 8.2 except \( \alpha_{PE} \), which was set to 0 dB SPL (see text for details).

**Figure 8.9.** Simulated intracellular receptor potential waveforms in response to 120 dB SPL tones of various frequencies. The tone frequency is indicated (in Hz) by the side of each curve. The amplitude scale is 35 mV between ticks. The model was evaluated with the parameters given in Table 8.2. The resting potential was \(-36.7\) mV in this case.

**Figure 8.10.** A comparison between experimental and simulated RI curves for stimuli at BF. (a) Representative RI curves of HSR, MSR and LSR fibres [adapted from figure 2 of Winter et al. (1990)]. The BF of these units varies from 16 and 24 kHz. (b) Simulated RI curves of LSR, MSR and HSR fibres. The model results were obtained for an 18 kHz BF, with the parameters shown in Tables 8.2 and 8.3 (see text for details). (c) Input-output function of the DRNL filter bank model with the same parameters. The thin dotted line illustrates linear behaviour. (d) Simulated AC and DC components of the receptor potential as a function of stimulus intensity for the same parameter set.

**Figure 8.11.** The effect of varying the parameters \( h \) and \( z \) of the transmitter release function individually. All the other parameters of the model have been kept constant. Notice how the SR decreases by decreasing \( z \), and how \( h \) controls the slope of the RI curve without changing the rate threshold substantially.

**Figure 8.12.** Spontaneous rate in the model as a function of the parameter \( z \) of the exponential transmitter release function.

**Figure 8.13.** (a) Modelled RI curves at 4 kHz BF using the same parameters set as for 18 kHz BF (Tables 8.2 and 8.3). (b) Modelled RI curves at 4 kHz BF for a smaller \( h \) (see Table 8.4). (c) DRNL I/O function at 4 kHz BF with the parameters given in Table 8.2, for a stimulus at BF (d) AC and DC components of the receptor potential at 4 kHz BF for a stimulus at BF.

**Figure 8.14.** Experimental and modelled RI curves for a HSR AN fibre for stimuli at and off BF. The BF of the fibre is 16 kHz. The experimental data (symbols) have been adapted from figure 2 of Yates et al. (1990). The modelled RI curves are
represented by the lines (see text for details). The stimulus frequency for each curve is indicated in the legend. (a) For stimulus frequencies at and below BF. (b) For stimulus frequencies at and above BF.

Figure 8.15. Experimental and modelled RI curves for a MSR AN fibre for stimuli at and off BF. The BF of the fibre is 16 kHz. The experimental data (symbols) have been adapted from figure 3 of Yates et al. (1990). The modelled RI curves are represented by the lines (see text for details). The stimulus frequency for each curve is indicated in the legend.

Figure 8.16. (a) I/O functions of the DRNL filter at 16 kHz BF. Different symbols indicate different stimulus frequencies. (b) The modelled DC component of the receptor potential as a function of stimulus intensity. These results correspond to the model RI curves shown in Figs. 8.14 and 8.15.

CHAPTER 9

Figure 9.1. Spectrum of the head-related impulse responses considered (a) S1, 0° elevation, 0° azimuth; and (b) S2, -20° elevation, 0° azimuth. Notice that the spectrum is represented as a function of the log_{10} of the frequency (in Hz), rather than the frequency itself. The correspondence between these two variables is shown in panel (c).

Figure 9.2. Model results at 4 (black squares) and 16 kHz (white squares) BF for broad-band noise stimuli. (a) DRNL filter input-output function. The dotted lines illustrate a linear behaviour. (b)-(d) Onset RI curves: (b) HSR fibre, (c) MSR fibre and (d) LSR fibre. The intensity in the X-axis is rms (See text for details).

Figure 9.3. Simulated steady-state RI curves at 4 and 16 kHz BFs for broad-band noise stimuli. (a) HSR fibre; (b) MSR fibre and, (c) LSR fibre. The intensity in the X-axis is rms, (See text for details). Notice that the rate threshold is approximately 10 dB higher than for pure-tone stimuli (Figs. 8.14 and 8.15).
Figure 9.4. (a) Output profiles from the DRNL filter bank. The symbols represent the output profile for the stationary stimulus S2 (–20° elevation, 0° azimuth). The thick lines represent the output profile for pure broad-band noise. The rms stimulus intensity (dB SPL) is indicated by the numbers on the left of each line. Notice that the $\log_{10}(BF)$ is represented in the X-axis [the correspondence between the actual BF and the $\log_{10}(BF)$ is given in Fig. 9.1c]. (b) Spectrum of S2.

Figure 9.5. Simulated rate profiles for a HSR fibre in response to the stationary stimulus S2. (a) Onset response; calculated as the maximum response over the first 10 ms of the stimulus. (b) Steady-state response; calculated as the average response over last half of the stimulus duration. (See text for details).

Figure 9.6. Simulated AN rate profiles for LSR, MSR and HSR fibres in response to the stationary stimulus S2. (a) stimulus intensity 20 dB SPL (rms). (b) 30 dB SPL (rms); (c) 40 dB SPL (rms); (d) 60 dB SPL (rms); and (e) 80 dB SPL (rms).

Figure 9.7. (a) Maximum-rate (onset) profiles over the first 10 ms of the moving stimulus MS (i.e., for the first source location). (b) Maximum-rate (onset-like) profiles over the period 50-60 ms of the moving stimulus MS (i.e., as the stimulus switches from the first to the second location). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms).

Figure 9.8. (a) Average-discharge rate-profiles over the first 50 ms of the moving stimulus MS (i.e., for the first source location). (b) Average-discharge rate-profiles over the last 50 ms of the moving stimulus MS (i.e., for the second source location). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms).

Figure 9.9. (a) The ratio of the magnitude spectra of the two stimulus locations computed as $20\log_{10}|S2/S1|$. (b) Maximum-rate (onset-like) profiles over the period 50-60 ms of the moving stimulus MS (i.e., as the stimulus switches from the first to the second HRTFs). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms). This figure is a reproduction of Fig. 9.7b.
CHAPTER 5

Table 5.1. Comparison between the central frequency of the experimental (Exp.) notches observed in the transfer function of the metallic spiral-shaped system (see Fig. 5.2) and the predictions of the single-delay-and-add (SDAA) approximation and the diffraction/reflection (DR) model. All values are expressed in kHz. The squared difference ($d^2$) between the experimental values and the predictions of both models is also shown. For every angle of elevation the sum of the squared differences is regarded as an indication of the overall performance of each model.

CHAPTER 7

Table 7.1. Parameters for the mechanical to neural transduction model (Meddis, 1986, 1988; Meddis et al., 1990).

CHAPTER 8

Table 8.1. The DRNL-filter parameters optimised at BF's of approximately 300 Hz, 8 kHz and 18 kHz (from O'Mard and Meddis, 1996).

Table 8.2. Model parameters. The asterisks denote those parameters whose value is different from the value proposed by the authors of the corresponding model.

Table 8.3. Set of parameters for the synaptic effects model (Meddis, 1986, 1988) and the transmitter release function. These parameters, together with those provided in Table 8.2, provide a complete set of parameters for simulating LSR, MSR and HSR fibres at 18 kHz BF. Note that only $h$ and $z$ vary across fibre types.
Table 8.4. Values of the parameters $h$ and $z$ for simulating LSR, MSR and HSR at 4 and 18 kHz BF. Notice that the values at 18 kHz coincide with those of Table 8.3. Notice that $z$ does not change with BF.

Table 8.5. Parameters for fitting the RI curves of MSR and HSR fibres shown in figures 2 and 3 of Yates et al., (1990). The other model parameters are the same as those given in Tables 8.2 and 8.3. Notice that only $h$ and $z$ vary for the two fibre types.

CHAPTER 9

Table 9.1. Model parameters.
CHAPTER 1

Introduction

1.1 MOTIVATION

What mechanisms enable us to localise sounds? When we hear a sound, what type of acoustic information is available to us about the position of the sound source? How is that information generated? Is the information the same for every individual? Can we actually use all the location-encoding acoustic information that is provided to us? It is hoped that by investigating the physical mechanisms underlying the generation of location-dependent spectral information in the pinna, and the processing of that information by the peripheral auditory system some insight into the way in which humans address the problem of sound localisation might be gained.

Sound localisation is commonplace in everyday life. Information of various kinds, not only of an acoustic nature, is provided to the central nervous system through our senses about the location of a particular sound source. The central nervous system is then able to combine all that information and judge the source location often precisely. Perhaps, the fact that the judgement is based upon a combination of cues (mainly visual, motional and acoustic) explains such a degree of accuracy.

In the laboratory, however, the amount of information provided to the central nervous system can be controlled so that only acoustic cues are available. Sound localisation is still possible in this case, although with considerably less accuracy. In fact, sound localisation accuracy depends on the type of stimulus and the experimental conditions. In general, in a free-field (anechoic) environment, we seem to be better at localising sounds with a broad spectrum than narrow-band or tonal stimuli (see Chapter 2,
Background. This suggests that important information about the source location is provided to the central nervous system in the spectral content of the stimulus.

The spectral content of a sound at the eardrum depends largely on the source location as a result of the directional-filtering effect of the human body. Of all the anatomical features of the body, the external ear, specifically the pinna, is mostly responsible for the generation of location-dependent spectral features. But, do pinna-based spectral features depend on source location in an orderly manner? If so, how does this dependence come about?, what are the physical mechanisms underlying the generation of pinna-based spectral features? More importantly, is their dependence generalisable to all pinna shapes? And can pinna cues be predicted for a given pinna shape?

Equally interesting is the issue of whether pinna-based spectral cues are fully signalled to the central nervous system. Even though location-dependent spectral information may be present at the eardrum, this does not necessarily mean that the information is available to the central nervous system. The transduction process of the acoustic stimulus from tympanic vibrations to auditory nerve discharge activity is nonlinear. As a result, the auditory-nerve (AN) representation of stimulus spectrum is likely to be a distorted version of the spectral shape measured at the eardrum. But, to what extent and under which conditions does this happen? What are the physiological limiting factors? Can we speculate or demonstrate the existence of any mechanism employed by the auditory system to extract pinna-based spectral information from the distorted AN representations? How does the representation vary for stationary and moving sound sources?

Providing an answer to these questions motivated the work presented in this thesis.

1.2 OBJECTIVES

Understanding and modelling both the physical mechanisms pertaining to the generation of pinna-based spectral cues and the physiological coding of those cues by the peripheral auditory system form the core of this thesis.
With regard to the first issue, namely the physical origin of pinna-based spectral cues, the following objectives were established: (A) to identify 'universal' pinna-based spectral features observed in human head-related transfer functions (HRTFs); (B) to characterise their dependence on source location (azimuth and elevation angles); (C) to investigate a possible one-to-one relationship between characteristic pinna-based spectral cues and individual anatomical components of the pinna; finally, and most importantly, (D) to model the transfer function of the pinna from first physical principles.

Because of the extremely complex structure of the pinna and its relatively-unknown acoustic characteristics, the aims of this last objective (D) are far-reaching. For this reason, our study is more restricted and focused only on providing a physical model of acoustic diffraction, reflections and interference phenomena in the main cavity of the pinna: the concha (Fig. 2.4). The formulation of such a model of the human concha was found to be invaluable in elucidating the physical mechanisms responsible for the characteristic location-dependency of some pinna-based spectral features, especially elevation-dependent notches observed for sources in the median and the lateral vertical planes (Hebrank and Wright, 1974; Shaw and Teranishi, 1968).

The effect of other anatomical features of the external ear (namely, the pinna-flange, tragus, crus helias, fossa, etc.; Fig 2.4) is investigated experimentally rather than theoretically, through objectives (A) to (C). The successful accomplishment of these three objectives requires the availability of a flexible, accurate experimental method for measuring head-related transfer functions which also guarantees the reproducibility of the measurements for later comparisons. Another aim of this work is to develop such a robust method. It was found that a method based on the use of a KEMAR (Knowles Electronics Manikin for Acoustic Research) fitted with individualised moulded-pinnae as the 'subject' for our HRTF measurements fulfilled the above requirements (Burkhard and Sachs, 1975). The use of a KEMAR was particularly advantageous with regard to objective (C), as the investigation on possible one-to-one relationships between pinna-based spectral features and specific anatomical features of the pinna was done by measuring the effects on head-related transfer functions of removing such anatomical features from the complete structure of the pinna. Obviously, such an approach would not have been possible using real people as subjects.
With regard to the physiological coding of pinna-based spectral cues, a major objective was the investigation, development and testing of a novel, exhaustive model of the peripheral auditory system which includes nonlinear cochlear filtering as a crucial characteristic. An important goal was to reflect realistically in this model all physiological processes and parameters that are known to affect, directly or indirectly, the pattern of auditory nerve activity in response to any stimulus. Such a model embodies our current understanding of the peripheral auditory processes in a fairly concise way and can form a firm basis for the investigation of processes at higher levels in the auditory system.

In this thesis, the proposed model is used in an investigation of the AN representation of HRTF-filtered broad-band stimuli. The following objectives were set with regard to this investigation: (E) to determine the extent to which pinna-based spectral cues are actually signalled to the central nervous system (CNS) at low, moderate and high signal levels; (F) to identify the limiting factors in the transduction process that determine the extent to which pinna-based spectral cues are signalled to the CNS giving special attention to how the nonlinear character of cochlear filtering affects the AN representation of the stimulus spectrum; (G) to test some hypothetical mechanisms that may be employed by the auditory system for extracting the spectral characteristics of the stimulus from the AN representation. With regard to this last objective, particular attention will be paid to investigating the transient AN representation of the stimulus spectrum as the stimulus moves from one location to another.

The benefits of studying these issues by modelling rather than experimentally are many (Hewitt, 1992). Perhaps the most interesting one in our case is the fact that the AN representation can be investigated in response to human HRTF stimuli. Only in a few instances (Poon and Brugge, 1993a,b; Rice et al., 1995) has the AN representation of pinna-based cues been investigated experimentally. In those cases, the study was done in cats. For obvious reasons, a similar type of experimental investigation is not possible in humans. Computer modelling provides an attractive alternative to experimentation in this case.

Far from intending to account for all the issues involving the generation and coding of pinna-based spectral cues, this study has a further goal: to lay the ground work and provide the tools for further more detailed investigations into those and other aspects
of sound localisation. For instance, it is of particular importance to study the benefits induced by a moving source and head movements concerning the amount of location-dependent information encoded in the AN response to these stimuli. It is equally interesting to investigate the benefits that a left-right comparison (or difference), or a serial differential mechanism could report to the CNS in order to disambiguate the location-dependent spectral information provided monaurally. The ultimate aim is to develop a realistic unified theory of sound localisation based upon feasible mechanisms available to the auditory system. In this sense, computer modelling provides a very useful flexible platform for making progress toward such a theory.

1.3 ORIGINAL CONTRIBUTIONS

The principal contributions of this work are:

- The characterisation of the dependence of pinna-based spectral features on source location; in particular, the azimuthal dependence of elevation-varying spectral notches (Chapters 3 and 4).

- The development of a physical model of sound diffraction, reflection and interference in the human concha (Chapter 5). The originality of this model lies in the fact that (1) it is the first attempt to our knowledge to simulate the transfer function of the human concha from first physical principles, and (2) it specifically includes sound diffraction within the concha cavity in its formulation. A paper that presents this model and some of the experimental results showed in Chapters 3 and 4 has been accepted for publication in The Journal of the Acoustical Society of America under the title "A Physical Model of Sound Diffraction and Reflections in the Human Concha."

- The characterisation of the contribution of individual anatomical features, namely the fossa cavity, the flange, the tragus and the crus helias (Fig. 2.4), to the overall head-related transfer functions (Chapter 6).

- The development and testing of a novel, exhaustive computer model of the peripheral auditory system which brings together some existing models of specific peripheral processes with new ones (Chapter 8). The originality of this model lies

---

mainly in the fact that it includes (1) nonlinear cochlear filtering and related phenomena (e.g., distortion products and two-tone suppression) (O'Mard and Meddis, 1996), and (2) an exponential transmitter release function which permits the modelling of low-, medium- and high-spontaneous rate AN fibres by changing only two parameters that characterise possibly the size of the IHC/AN synaptic plaque and body.

- The identification (through modelling) of the factors that determine the quality of AN rate representation of pinna-based spectral cues at low, moderate and high signal levels (Chapters 7 and 9).

In the process of developing these ideas, further contributions have been made, some of which are:

- The validation and documentation of a technique for making individualised moulded pinnae to be fitted into a KEMAR (Chapter 2).
- The improvement of a experimental technique for measuring head-related transfer functions in real time using a KEMAR, and the adaptation of the necessary software so that it can be run in a Silicon Graphics (Indy) work-station (Chapters 3 and 6). With the improved technique, HRTF measurements can be made even in reverberant environments (Chapter 6).
- The measurement and analysis of the head-shadow effect associated with the head of a KEMAR for a large number of source locations (Chapter 3).
- A further insight into possible morphological and functional differences at the level of the IHC/AN synapse which might explain the existence of low-, medium-, and high-spontaneous rate AN fibres (Chapter 8).
- The characterisation of a kind of differencing mechanism observed in transient response of the AN to changes in the spectral characteristics of the stimulus. This differencing mechanism is related to the adaptation characteristics of AN fibres and may be a useful way for enhancing the contrast of pinna-based spectral features under certain conditions (Chapter 9).
- An examination of the correlation between reported psychophysical results and the 'physiological' predictions of our model with regard to sound localisation tasks where pinna-based spectral features are the only available cue.
Finally, although the scope of this thesis does not include physiological or psychophysical experimentation, a discussion on the physiological and psychophysical implications of the findings is provided in many cases. In some instances physiological and psychophysical experiments are suggested to test the model predictions.

1.4 OVERVIEW OF THE THESIS

The thesis begins with a Background chapter (Chapter 2). This chapter serves two main purposes: (1) to provide a general, integrated framework to the thesis, and (2) to present the objectives formulated above in the context of the current knowledge. Additionally, it attempts to point out some important issues on sound localisation which are usually unnoticed.

Chapter 3 covers several issues, all of them inter-related. Firstly, a technique for making pinna casts from ears of real people (or individualised moulded pinna) so that they can be fitted into a KEMAR is presented. Secondly, an experimental method for measuring head-related transfer functions is described. Thirdly, experimental head-related transfer functions obtained with this method are presented for three different pairs of individualised moulded pinnae and a large number of azimuthal angles. The results are compared with published data in order to (a) validate the proposed technique of pinna-moulding and the method for measuring HRTFs, (b) identify general spectral features observed across subjects, and (c) characterise the azimuthal dependency of those features, especially notches. Finally, HRTF measurements obtained with the manikin wearing no pinnae are shown. These measurements were made to characterise the effects of the torso and the head-shadow effect.

In Chapter 4, the elevation dependency of spectral features observed in HRTFs is investigated. Special attention is given to characterise the elevation dependency of spectral notches for sources in vertical planes other than the median and the lateral vertical planes (these planes are defined below in the section Angle conventions). Characteristic elevation-varying spectral notches have been previously observed for sources in the median vertical plane (Hebrank and Wright, 1974). These notches are believed to be responsible for the sensation of elevation (Hebrank and Wright, 1974;
Butler and Belendiuk, 1977). In Chapter 4, it is investigated whether similar notches show consistently the same elevation-dependency behaviour for sources in other vertical planes. If that were the case, this would not only suggest that the speculated psychophysical role of these notches could be extended from the median plane to any other vertical plane, but also it would indicate that a common physical mechanism underlies the generation of these elevation-dependent notches whatever the azimuthal location of the sound source.

The conclusions drawn from the experimental HRTFs shown in Chapters 3 and 4, together with some speculative theories of concha reflections proposed by Hebrank and Wright (1974) led to the work presented in Chapter 5. In this chapter, a physical model of sound diffraction, reflection and interference in the concha cavity of the human pinna is developed and tested. The model attempts to account for the characteristic elevation-dependent spectral notches described in Chapters 3 and 4. The role of sound diffraction as the sound waves enter the concha cavity is specifically examined, as it is hypothesised that sound diffraction is responsible for the azimuthal dependency of the notches. The model is tested for an acoustic diffracting and reflecting system with a simple spiral shape, before being applied to a realistic concha shape geometry. Because some approximations are made in order to simplify the problem, the proposed model must be regarded as an approximation. However, much knowledge about the directional-filtering nature of the concha (and the pinna in general) is gained from such a model. Ways forward to find an exact solution are suggested.

Chapter 6 is concerned with examining the contribution of individual anatomical features of the pinna to the overall HRTF. This is done by measuring and analysing the effect that gradually removing these anatomical features has on the overall HRTF. An improved version of the experimental method described in Chapter 3 is also presented in this chapter. It is explained that amongst other advantages, the improved method is much quicker and permits the measurements to be carried out in a reverberant environment. The effect of occluding the flange cavities (fossa and scapha), and gradually removing the whole flange, the crus helias and the tragus (Fig. 2.4) is measured with this improved method for a large number of source locations. The idea behind this type of analysis is to investigate possible one-to-one relationships between characteristic spectral features observed in HRTFs and these anatomical features of the pinna. Because of the complexity of the acoustic phenomena involved
in pinna filtering, such a relationship could not be specifically found in many cases. However, a particularly important finding was that the flange cavities play an important role in determining the spectral characteristics observed in HRTFs, which suggests that such cavities should be included in a future, further-developed physical model of the pinna.

In Chapter 7, the AN rate representation of HRTF-filtered stimuli is investigated using computer simulation. A computer model that simulates the activity of low-, medium- and high-spontaneous rate (LSR, MSR and HSR) AN fibres is presented. The model includes a linear gamma-tone auditory filter bank to simulate the cochlear filtering. The proposed 'linear' model is used mainly to investigate (1) whether the AN rate representation of HRTF-filtered stimuli conveys more spectral information at the stimulus onset than in the steady-state, and (2) the extent to which the rate threshold and the dynamic range of different types AN fibres determine the rate representation of the stimulus spectrum. Because this model only includes a linear approximation of the nonlinear cochlear filtering, it is inadequate for investigating the effects that cochlear nonlinearities and related phenomena (two-tone suppression and distortion products) have on the AN rate representation of the stimulus spectrum. However, the work presented in this chapter served as a training exercise and a pilot study for anticipating the problems and determining the aims to be achieved with a much more realistic and novel model of the peripheral auditory system which is described in Chapter 8.

Chapter 8 presents a novel, functional model of the peripheral auditory system which includes a dual-resonance nonlinear (DRNL) auditory filter bank model (developed by O'Mard and Meddis, 1996). An attempt is made to embody in this model the current knowledge on the processes and parameters that determine, in one way or another, the rate activity of AN fibres. The model includes some published models of specific peripheral processes such as the DRNL filter bank by O'Mard and Meddis (1996), a model (by Shamma et al., 1986) of the relationship between the basilar membrane (BM) motion and the displacement of the inner hair-cell (IHC) stereocilia, a model of the IHC intracellular receptor potential (Shamma et al., 1986), and a new, stochastic version of the original model of the synaptic effects by Meddis (1986, 1988). In addition to these published models, a novel transmitter-release exponential function is proposed based on physiological evidence (Stevens, 1968; Sewell, 1984). The significance of an exponential function is examined with special care as it is found to be very appropriate for simulating the rate activity of LSR, MSR and HSR AN fibres,
and its two associated parameters may account for the morphological differences between the three fibre types. The model parameters are optimised to simulate experimental data at the various stages in the transduction process (cochlear filtering, IHC receptor potential and AN rate-intensity curves), for stimulus frequencies at- and off- the fibre's best-frequency (BF). Although the model can work in principle at BFs below 4 kHz, the parameters are optimised to obtain a realistic response at BFs across the range 4 to 17 kHz, as this is the audible range of pinna-based spectral features observed in HRTFs.

The computer model of the peripheral auditory system developed in Chapter 8 is used in Chapter 9 for investigating the AN rate representation of pinna-based spectral cues for stationary and moving HRTF-filtered broad-band noise stimuli. Special efforts are made to find out the extent to which the nonlinear cochlear filtering determines the quality of the rate representation of the stimulus spectrum for stationary stimuli at high signal levels. Additionally, the investigation carried out in Chapter 7 with the linear gamma-tone model about the influence of the fibre's rate characteristics on the AN rate representation of the stimulus spectrum is repeated here with the nonlinear model presented in Chapter 8. The transient AN rate representation associated with a stimulus switching from one location (or HRTF) to another is also investigated with the nonlinear model. A differencing mechanism that enhances the contrast in the AN representation of the stimulus spectrum is presented. This mechanism is characterised with the model and a tentative explanation is provided in terms of the adaptation characteristics of the fibres and their response to step-like functions. Some of the findings presented in this chapter and their psychophysical implications are subject to confirmation by experimentation. However, they are discussed in the context of current physiological and psychological knowledge.

In the final chapter (Chapter 10), the work covered in this thesis is summarised, and particular strengths and weaknesses identified. The way forward is assessed and a number of ideas on approaching some outstanding issues are presented. There remain, as ever, many problems to be solved in the pursuit of a complete understanding of the issues involved in the physical origin and physiological coding of pinna-based spectral cues, but it is hoped that this work will help to provide some further insights into those issues.
1.5 ANGLE CONVENTIONS

In order to avoid ambiguity, the following angle conventions have been adopted throughout this thesis:

- **Elevation:** Elevation angles ($\phi$) have values within the range $-90^\circ \leq \phi \leq +90^\circ$, so that $+90^\circ$ corresponds to a source immediately above the head, $0^\circ$ corresponds to a source at eye level, and $-90^\circ$ corresponds to a source immediately below the feet.

- **Azimuth:** Azimuthal angles ($\theta$) have values within the range $0^\circ \leq \theta < 360^\circ$, measured clockwise from the right-ahead position, so that for a source at $0^\circ$ elevation, $\theta = 0^\circ$ corresponds to a source right ahead, $\theta = 90^\circ$ corresponds to a source directly to the right ear, $\theta = 180^\circ$ corresponds to a source behind the head, and $\theta = 270^\circ$ corresponds to a source directly to the left ear.

The following planes are frequently referred to in this thesis:

- **Horizontal Plane** (HP): This is the plane at eye-level ($0^\circ$ elevation), through the ears, that separates the top and the bottom hemi-fields.

- **Median Vertical Plane** (MVP): This is the plane of 'symmetry' of the body; that is, the plane that separates the left and the right hemi-fields.

- **Lateral Vertical Plane** (LVP): This is the plane through the ears that separates the front and the back hemi-fields.

- **Ipsilateral Diagonal Vertical Plane:** This is the vertical plane at $45^\circ$ from the lateral vertical plane towards the front of the subject.

- **Contralateral Diagonal Vertical Plane:** This is the vertical plane at $135^\circ$ from the lateral vertical plane towards the front of the subject.
CHAPTER 2

Background

2.1 INTRODUCTION

This chapter has two purposes: (a) to provide a general, integrated framework to the thesis, and (b) to present the objectives of this thesis (see Chapter 1) in the context of the current knowledge.

An appropriate place to begin is by explaining that when we hear a sound in a free-field (anechoic) situation, the pressure field at each eardrum, $e(t)$, is a filtered version of the original pressure variations generated by the sound source. The human body filters the incoming sounds before they reach the eardrum in a way which depends on the stimulus location. Mathematically, this can be expressed as follows:

$$ e(t) = a(t) \ast s(t) $$

(2.1)

where: $s(t)$ is the stimulus signal, $a(t)$ is the impulse response of the human body (known as the head-related impulse response, or HRIR), and $e(t)$ denotes the time signal at the eardrum calculated as the convolution of $a(t)$ and $s(t)$. Mathematically, $a(t)$ is defined as $e(t)$ for a Dirac-pulse stimulus (see Chapter 3). For simplicity, Eq. (2.1) is usually expressed in the frequency domain, so that the convolution product becomes a standard multiplication:

$$ E(f) = A(f)S(f) $$

(2.2)

where: $S(f)$, $A(f)$ and $E(f)$ are the Fourier-transforms of $s(t)$, $a(t)$ and $e(t)$ respectively. In this case, $A(f)$ usually represents the free-field-to-eardrum transfer function of the
human body, commonly known as the head-related transfer function, or HRTF (Wightman and Kistler, 1989). Therefore, $E(j)$ represents the gain at the eardrum for the stimulus $S(j)$ with respect to a point at the centre of head in a situation where the subject is not present (Blauert, 1983).

$A(j)$ depends on the direction of incidence of the stimulus. In other words, $A(j)$ is different for every stimulus location. Considering that the two ears are on opposite sides of the head, this means that the HRTFs for the left and the right ears, $A_L(j)$ and $A_R(j)$, are different for most source locations. As a result, the simultaneous pressure fields at the left and right eardrums, $e_L(t)$ and $e_R(t)$, are different for most source locations.

From a psychophysical point of view, $e_L(t)$ and $e_R(t)$ are all the information available to the auditory system in normal conditions. With this information, the auditory system has to identify and localise the sound source. Obviously, the information about the identity of the sound source is encoded in $s(t)$, whereas the information about the source location is encoded in $a_L(t)$ and $a_R(t)$, or, equivalently, $A_L(j)$ and $A_R(j)$.

As explained in Chapter 1, part of the work presented in this thesis is about providing a theoretical description and understanding of the physical mechanisms that generate the source-location information that is available to the auditory system. In other words, about explaining the physical processes that give rise to $A_L(j)$ and $A_R(j)$.

Explaining HRTFs by studying the transfer function of individual anatomical features

The problem of understanding and modelling the physical mechanisms that generate HRTFs is a complex one. A common approach to simplify the problem consists in measuring and modelling separately the individual contribution of the head, the torso, the pinna, and the meatus. The idea behind this type of analysis is that HRTFs can be then expressed directly in terms of the transfer functions of individual anatomical features. In this sense, it is common to see $A(f)$ treated (directly or indirectly) as:

$$A(f) = T(f)H(f)P(f)M(f)$$

(2.3)
(e.g., Shaw, 1975b) where, $T(f)$, $H(f)$, $P(f)$ and $M(f)$ are the frequency transfer functions of the torso, the head, the pinna and the meatus respectively. This approach is convenient, but it is incorrect since it does not consider the acoustic interactions between the various anatomical elements. Therefore, this form of analysis must be considered as an approximation, valid only at low frequencies (< 7 kHz) (Shaw, 1975b).

Despite this, much understanding about the location-dependent information encoded in HRTFs has been gained by adopting this approach. Very comprehensive reviews of the various studies on the transfer functions of individual anatomical features have been given by Shaw (1975b), Blauert (1983) and Kuhn (1987). For this reason, this chapter will attempt to review only some important general issues, stressing some points which are usually unnoticed.

### 2.2 SOURCE-LOCATION INFORMATION ENCODED IN BINAURAL HRTFs

As explained above, the directional filtering of the human body has an important consequence: for a given sound source position, the simultaneous pressure fields at the left and right eardrums, $a_L(t)$ and $a_R(t)$, are different. In other words, the HRTFs for the left and right ears, $A_L(f)$ and $A_R(f)$, corresponding to a particular source location are different. This statement is true for most source locations. This means that not only do $a_L(t)$ and $a_R(t)$ vary individually with source position, but also the difference between them is a function of source position, which can be used by the auditory system as another localisation cue. Therefore, the source-location information encoded in binaural HRTFs can be divided in two groups according to the way in which it is presented to the auditory system: binaural information and monaural information.
2.2.1 Binaural information

The effect of the head: Interaural time and level differences

The fact that the two ears are situated on opposite sides of the head generates interaural differences in the time of arrival of the acoustic stimulus. Additionally, the presence of the head between the two ears generates interaural level differences by shadowing the sound at the ear on the far side from the stimulus. Lord Rayleigh (Rayleigh, 1907) is known to be the first person who proposed and tested a theory of sound localisation based upon the existence of these two binaural cues. He concluded that the interaural level difference (ILD) must be the operational localisation cue for high frequency stimuli, whereas the interaural time difference (ITD) must be the operational localisation cue for low frequency stimuli. The trade-off frequency between the two mechanisms is around 1000 Hz. These conclusions were based upon experimental evidence and the fact that low frequency sounds will diffract more easily around the head without losing intensity, which means that ILDs are negligible for low frequency stimuli. The notion that spatial information is derived at high frequencies from head-based ILDs and at low frequencies from head-based ITDs is often referred to as the "duplex" theory of sound localisation.

The spatial information encoded by ITDs and ILDs is intrinsically present in binaural HRIRs. In studying and modelling head-based ITDs and ILDs analytically, it is common to approximate the head by a perfect sphere of similar dimensions, with the two ears regarded as points on opposite sides of the sphere. Precise mathematical descriptions of the pressure field at any point on the surface of a rigid sphere have been given by Lord Rayleigh (Rayleigh, 1904), Ballantine (1928) and Rschevkin (1963). However, considerable knowledge about the dependence of ITDs on source location can be gained from simple geometrical descriptions of the interaural path difference for a sound source outside the head. An excellent review of the various models proposed so far has been given by Blauert (1983, p. 70 ff.).

A simple geometrical model of ITDs

A simple geometrical model of ITDs can be constructed from Fig. 2.1. The head is approximated by a perfect sphere (of radius $a$) and the ears are regarded as points on both sides of the head ($L$ and $R$ for the left and right ears respectively). The sound
source is indicated by \( P(x_p, y_p, z_p) \) in Cartesian co-ordinates or, equivalently, \( P(d, \theta, \phi) \) in spherical co-ordinates, where: \( d \) is the distance from the sound source to the centre of the sphere, \( \theta \) is the azimuthal angle, and \( \phi \) is the elevation angle defined as in Fig. 2.1. The interaural time difference is calculated as \( \text{ITD} = (D_L - D_R)/c \), where \( D_L \) and \( D_R \) are the distances \( \overline{PL} \) and \( \overline{PR} \) from the sound source to the left and the right ears respectively, and \( c \) is the speed of sound.

From the 3D geometrical description given in Fig. 2.1a, \( D_L \) and \( D_R \) can be calculated as follows:

\[
D_L = \overline{PQ} + \arccos(QL) = \sqrt{d^2 - a^2} + a \gamma_x = \sqrt{d^2 - a^2} + a(\pi - \gamma_1 - \gamma_2) \quad (2.4)
\]

\[
D_R = \overline{PR} = \sqrt{C^2 + (x_p - a)^2} = \sqrt{(y_p^2 + z_p^2) + (x_p - a)^2} = \sqrt{d^2 + a^2 - 2ax_p} \quad (2.5)
\]

Therefore, the resultant ITD according to this 3D geometrical model is given by:

\[
\text{ITD} = \frac{1}{c} (D_L - D_R) = \frac{1}{c} \left[ \sqrt{d^2 - a^2} + a(\pi - \gamma_1 - \gamma_2) - \sqrt{d^2 + a^2 - 2ax_p} \right] \quad (2.6)
\]

where \( \gamma_1 \) and \( \gamma_2 \) can be also calculated from Fig. 2.1 as:

\[
\gamma_1 = \tan \left( \frac{\overline{PQ}}{a} \right) = \tan \left( \frac{\sqrt{d^2 - a^2}}{a} \right) \quad (2.7)
\]

\[
\gamma_2 = \tan \left( \frac{C}{x_p} \right) = \tan \left( \frac{\sqrt{y_p^2 + z_p^2}}{x_p} \right) \quad (2.8)
\]

The ITD can be expressed in terms of the source distance, and the azimuthal and elevation angles by converting Eq. 2.6 from Cartesian to spherical co-ordinates with the following relationships:

\[
x_p = d \cos \phi \sin \theta \quad (2.9a)
\]

\[
y_p = d \cos \phi \cos \theta \quad (2.9b)
\]

\[
z_p = d \sin \phi \quad (2.9c)
\]
Figure 2.1. (a) 3D Geometrical model of ITDs. The head is approximated by a perfect sphere and the ears are regarded as points on both sides of the head (L and R for the left and right ears respectively). The sound source is indicated by $P(x_p, y_p, z_p)$ in Cartesian co-ordinates or $P(d, \theta, \phi)$ in spherical co-ordinates. (b) 2D geometrical model of ITDs. Notice that this representation is equivalent to (a) but with $\phi=0$. In both models, the interaural time difference is calculated as $\text{ITD} = (D_L - D_R)/c$. 
The results of this model are shown in Fig. 2.2. The model was evaluated for a sphere of radius 9.3 cm (Kuhn, 1977), and $c = 340$ m/s. Fig. 2.2a shows the ITD as a function of the azimuthal angle of the source ($\theta$). Results are presented at elevation angles ($\phi$) from $0^\circ$ to $90^\circ$, in $10^\circ$-steps. Fig. 2.2b shows the ITD as a function of source elevation. In this case, the azimuthal angle is the parameter which varies from $0^\circ$ to $90^\circ$, in $10^\circ$-steps. The results presented in Figs. 2.2a and 2.2b were obtained for a source at a distance of 1 m from the centre of the sphere. Fig. 2.2c shows the dependence of ITD on source distance for a source at $0^\circ$ elevation. The parameter is the azimuthal angle, which varies from $0^\circ$ to $90^\circ$ in $15^\circ$-steps.

Because of the clear dependence of the ITDs on source azimuth—particularly at $0^\circ$ elevation—(see Fig. 2.2a), ITDs are usually considered as a useful localisation cue only for sources in the horizontal plane (for a complete review see Middlebrooks and Green, 1991). It is usually unnoticed that ITDs can be also used for estimating source elevation. As shown in Figs. 2.2a and 2.2b, ITDs encode as much information about source elevation for sources at $90^\circ$ azimuth (lateral vertical plane), as about source azimuth at $0^\circ$ elevation (horizontal plane). Another important result from this simple model is that ITDs also encode some information about source distance (Fig. 2.2c), particularly for sources close to the surface of the head ($d < 30$ cm).

Figure 2.2 (shown in the next page). ITDs for a spherical head (of radius $a = 9.3$ cm) as predicted by a simple geometrical model (see text for details). (a) ITD as a function of the azimuthal angle, for 10 different elevations between $0^\circ$ and $90^\circ$ in steps of $10^\circ$. (b) ITD as a function of the elevation angle, for 10 different azimuths between $0^\circ$ and $90^\circ$ in steps of $10^\circ$. The results presented in (a) and (b) are for a source 1 m away from the centre of the sphere. (c) ITD as a function of source distance, for azimuths between $0^\circ$ to $90^\circ$ in steps of $15^\circ$. Distances are measured from the centre of the sphere.
Figure 2.2. See legend in the previous page.
The pressure field at any point on a rigid sphere

The amount of information about interaural differences that can be obtained from simple geometrical models is very limited. For instance, geometrical models can not account for the dependence of ITDs on the stimulus frequency (Kuhn, 1977). Most importantly, geometrical models of the difference in the path from a sound source to the two ears can not simulate ILDs.

For this reason, a complete description of the pressure field at any point on the surface of a sphere is required in order to model ITDs and ILDs properly. Such a description was first given by Lord Rayleigh (1904). Later solutions to the problem have been given by Ballantine (1928) and Rschevkin (1963). The solutions by Ballantine [Eq. (2.10)] and Rschevkin [Eq. (2.11)] for the pressure gain at a point on a rigid sphere (of radius $a$) at an angle $\alpha$ from the direction of incidence of the sound are:

\[
\left( \frac{p_i + p_s}{p_o} \right)_{r=a} = \left( \frac{2}{z\pi} \right)^{\frac{1}{2}} \exp(-iz) \sum_{m=0}^{\infty} \left( -1 \right)^m \frac{i^m (2m+1) P_m(\cos\alpha)}{(m+1) J_{m-\frac{1}{2}} + z J_{m-\frac{1}{2}}} \frac{i [(m+1) J_{m+\frac{1}{2}} - z J_{m+\frac{1}{2}}]}{[(m+1) J_{m+\frac{1}{2}} - z J_{m+\frac{1}{2}}]} \quad (2.10)
\]

\[
\left( \frac{p_i + p_s}{p_o} \right)_{r=a} = \sum_{m=0}^{\infty} i^{m+1} (2m+1) P_m(\cos\alpha) \left[ \frac{n_m j_m - j_m n_m}{j_m - in_m} \right] \quad (2.11)
\]

where: $p_i$ is the incident pressure at the observation point on the surface of the sphere; $p_s$ is the scattered pressure at the same point; $p_o$ is the pressure that would be at the same point if the sphere was not present; $i = \sqrt{-1}$; $z = ka$, with $k = 2\pi/\lambda$, $\lambda$ being the wavelength; $P_m$ is the Legendre polynomial of $m$-th order, $J_m(z)$ is the Bessel function of $m$-th order and argument $z$, $j_m(z)$ is the spherical Bessel function of $m$-th order and argument $z$, and $n_m(z)$ is the spherical Neumann function of $m$-th order

1 Although Eqs. (2.10) and (2.11) look different, their physical meaning is identical and it can be demonstrated that they are mathematically equivalent. However, Eq. (2.10) is less expensive in terms of computational time. Both equations converge very slowly; therefore, a large number of terms is required in the summation (i.e., a large $m$) to obtain correct results. The minimum value of $m$ is determined by the frequency. For frequencies up to 15 kHz, $m$ should be at least 50.
and argument $z$ (the definition of these functions can be found in most advanced books on acoustics, for instance, Morse and Ingard, 1968).

Eqs. (2.10) and (2.11) allow us to simulate the head-shadow effect as well as to calculate ILDs and ITDs.

**Head-shadow effect.** Fig. 2.3 shows the intensity gain\(^2\) (i.e., the head-shadow effect) after evaluating Eq. (2.11) at five different angles $\alpha = 0^\circ, 50^\circ, 90^\circ, 135^\circ$ and $180^\circ$, for a sphere of radius 8.75 cm. For the case that we are interested in calculating the pressure gain at the left ear for a sound source at $0^\circ$ elevation, these angles correspond respectively to the following azimuthal angles: $270^\circ, 320^\circ, 0^\circ, 45^\circ$ and $90^\circ$ (see inset in bottom-right corner of Fig. 2.3). [Culling et al. (1994) have published the results of evaluating Eq. (2.11) for a larger number of angles.] It is worth pointing out some of the results shown in Fig. 2.3:

1. The head-shadow effect is greater for higher frequencies. This reflects the fact that high frequencies do not diffract easily around the head (Rayleigh, 1904), and that objects much smaller than the wavelength have little influence on a sound.
2. For an incident sound straight on the left ear ($\alpha = 0^\circ, \theta = 270^\circ$), the presence of the head generates a 6 dB gain for frequencies above 5 kHz.
3. There is hardly any difference in the attenuation introduced by the head at $\theta = 270^\circ$ and $\theta = 320^\circ$, as in both cases the source is on the ipsilateral side.
4. The maximum head-shadow effect does not occur for a sound incident right on the opposite side of the head (i.e., at an angle $\alpha=180^\circ$ from our measuring point) as might be expected. Fig. 2.3 shows that the attenuation is even larger at an angle $\alpha = 135^\circ$, i.e., at an azimuth angle $\theta = 45^\circ$. As a matter of fact, results published by Culling *et al*. (1994) show that maximum attenuation occurs approximately at an angle $\alpha = 160^\circ$ with respect to our measuring point.

\(^2\) The intensity gain is expressed in dB and was calculated as $20\log_{10}[\|p_i + p_s\|/P_0]$. 
Figure 2.3. Intensity gain (in dB) at the left-ear point in a spherical head. The values were calculated as $20 \log_{10} \left( \frac{p_i + p_s}{p_o} \right)$, where $(p_i + p_s)/p_o$ was obtained from evaluating Eq. (2.11), (see text for details). $\theta$ represents the azimuthal angle and $\alpha$ represents the angle between the incident direction of the sound and the point at which the pressure gain is being measured (see inset on bottom-right corner).
**Head-based ILDs.** Head-based ILDs can be simulated as the difference between the results of evaluating either Eq. (2.10) or Eq. (2.11) for angles $\alpha_L$ and $\alpha_R$ (notice that $\alpha_L + \alpha_R = 180^\circ$) corresponding to the left- and right-ear points for a given direction of incidence of the sound. Blauert (1983) and Kuhn (1987) have shown the resultant ILDs as a function of source azimuth and frequency. In Blauert's results, ILDs as large as 20 dB SPL are predicted for some azimuths, which agrees with the experimental values measured by Middlebrooks et al. (1989).

**ITDs.** Eqs. (2.10) and (2.11) can be also used to calculate interaural phase differences (IPDs), from which ITDs can be derived as $\text{ITD} = \text{IPD}/\omega$, with $\omega = 2\pi f$. Kuhn (1977, 1987) validated Eq. (2.11) by comparing its predictions regarding ITDs with corresponding experimental data. He found that ITDs depend not only on source azimuth but also on the stimulus frequency. For instance, for a sound source incident directly on one ear, the ITD varies from 0.8 ms for frequencies up to 500 Hz, to 0.65 ms for frequencies above around 2 kHz. Finally, with regard to the dependence of ITDs and ILDs with source distance, Hartley and Fry (1921) showed that the calculated ILDs depend heavily on source distance, whereas ITDs shows little or no such dependence (Blauert, 1983 p. 75).

**Limitations of the "Duplex" theory of sound localisation**

We have seen above that ITDs and ILDs generated by the mere presence of the head between the two ears encode some information about source location as a result of their dependence on azimuth, elevation and distance. However, head-based ITDs and ILDs are not enough by themselves to encode unambiguously the position of a sound source in space. As explained by Wallach (1940, p.340), "...they (head-based ITDs and ILDs) only determine how far from the median plane, on the left or on the right side, the given sound source is located. Whether it lies on the front or in the rear, above or below the horizontal plane, remains undetermined, and the same is true of the amount of its elevation." For instance, it can be seen in Figs. 2.2a and 2.2b that a given ITD corresponds to many different azimuthal and elevation angles. Another interesting case is that of sound localisation in the median vertical plane, where head-based interaural differences are not available assuming a perfectly symmetrical head. However, it is now known that sound localisation in these cases is possible to some extent (see
below), which indicates that localisation cues of a different nature must be provided to and used by the auditory system in these situations.

Wallach (1940) was one of the first to suggest that head movements and vestibular and visual cues help to solve the ambiguity generated by head-based ITDs and ILDs. However, sound localisation is also possible in situations where both the head and the sound source remain still (for a complete review on the psychophysical studies on sound localisation see Middlebrooks and Green, 1991; Blauert, 1983; Moore, 1989). In these cases, the only remaining source of directional cues is provided by the filtering action of the pinnae. Ample background to pinna-based directional cues is given in the next sections, as are they are the main object of study of this thesis.

**Pinna disparity: another source of binaural information in the form of ILDs**

In the next section, pinna-based directional cues will be presented as monaural cues. This is based on the assumption that the head and the pinnae are perfectly symmetrical. Of course, this is not really the case. Searle *et al.* (1975) demonstrated that pinna disparity is another source of binaural information. The slight differences between the shapes of the left and the right pinnae generate pinna-based ILDs, at least within narrow frequency bands, that can be as large as 15 dB for sources in the median vertical plane, where head-based ILDs are nil. According to Searle *et al.* (1975), this binaural information may be used by the auditory system in sound localisation tasks, particularly for sources in the median vertical plane where head-based interaural differences are not available.

Notice that a subtle distinction has just been made between head-based and pinna-based ILDs, in an attempt to stress the fact that binaural information is related to two different aspects of the filtering process by the human anatomy; namely the fact that the ears are on opposite sides of the head and the fact that the head and the two ears are not exactly symmetrical. It must be stressed, however, that both types of ILDs are present in binaural HRIRs but might be indistinguishable to the auditory system.
2.2.2 Monaural information: pinna-based spectral cues

As explained above, monaural HRIRs also encode information about the position of the sound source. This information is generated by the directional filtering effect of the external ear, particularly the pinna (Fig. 2.4).

Batteau (1967, 1968) was one of the first to point out the importance of the pinna in sound localisation. He suggested that its role is to introduce a series of delays (or echoes) which encode precise information about the location of the sound source, helping to resolve, therefore, the ambiguity associated with head-based interaural differences. In his view, pinna delays are generated by reflections of the incident sound on the folds and cavities of the pinna (Fig. 2.4).

Batteau's ideas on the importance of the pinna in sound localisation tasks were supported at that time by the results of previous studies conducted by Batteau and Plante (1962) from which it was concluded that sound localisation by persons totally deaf in one ear is commonly observed. Additionally, results by Freedman and Fisher (1968) also indicated that the pinnae, even someone else's pinnae, appear to be necessary for sound localisation. More recently, Gardner and Gardner (1973) have shown that sound localisation accuracy for sources in the median plane decreases with greater occlusion of the pinna cavities, which again strengthens the view that the pinna plays an important role in sound localisation.

The nature of the location-dependent information provided by the filtering action of the pinna: temporal or spectral cues?

The way in which the auditory system uses the location-dependent information provided by pinna filtering has been a matter of debate. Batteau's theory was enunciated in the time domain. That is, he suggested that the 'echoes' generated in the pinna can be detected and adequately interpreted by the auditory system to reveal the location of the sound source. Batteau and Plante (1962) measured the delay in the time of arrival between the direct and reflected sound from the pinna walls. They found that the delay varies monotonically from 10 to 100 µs for azimuth angles, and from 100 to 300 µs for elevation angles. Results from the experiments by Wright et al. (1974)
supported Batteau's theory, as they showed that delays as short as 20 μs can be detected by the auditory system.

On the other hand, the experiments by Thurlow and Runge (1967) showed that sound localisation for sources in the median plane was poorer for clicks than for white noise stimuli. Wright et al. (1974), discussed this result as follows: "...if the delay mechanism (proposed by Batteau) did exist, one would expect that steep wavefronts presented by clicks to be equally well, if not better, detected and correlated than noise peaks. Such experiments (i.e., those by Thurlow and Runge, 1967) lead one to speculate that only spectral characteristics of a composite signal are evaluated in the localisation process." In other words, the results by Thurlow and Runge (1967) indicate that location-dependent pinna cues must be interpreted by the auditory system as spectral cues rather than temporal cues. Further evidence in support of this view is given by Watkins (1978), who also proposed a model of sound localisation based upon spectral-pattern recognition, and Blauert (1969/70), who showed that the perceived source location is a function of the frequency content of the stimuli and does not depend on the actual location of the sound.

Because of the short delays associated with pinna reflections, the fact is that most current theories imply that the pinna produces changes in the spectrum\(^3\) of the incident source before it reaches the tympanic membrane (for a review see Middlebrooks and Green, 1991). This is also the interpretation adopted in this thesis. Under this interpretation, each source location has an associated 'spectral shape', which may be used as a localisation cue by the auditory system. As explained by Middlebrooks and Green "...patterns of spectral features associated with particular locations are referred to as 'spectral shape cues'. Because these cues are a product of the acoustics of the pinna, they often are referred to as 'pinna cues'," or pinna-based spectral cues. Notice that pinna-based spectral cues are often referred to as 'monaural cues', as they do not require interaural comparison to encode source location information (but see above, section Pinna disparity: another source of binaural information in the form of ILDs).

\(^3\) From a physical point of view, the delays generated by the pinna produce changes in the spectrum of the incident sound when the total sound entering the meatus is expressed as the sum of the direct plus the delayed waves (Wright et al., 1974).
Figure 2.4. The external ear (adapted from Shaw, 1975b). Notice the three major sub-divisions of the external ear: the meatus (dark pattern), the concha (light pattern) and the pinna-flange (non-shaded area).
Identification of pinna-based spectral cues in HRTFs

HRTFs represent the transformation of sound pressure from the free field to the eardrum, regarding the human body as a whole. However, this transformation is largely determined by acoustic reflections, diffraction, scattering, resonance and interference effects in individual anatomical structures such as the torso, the head, the pinnae and the meatus. To some extent (see above), particular spectral features observed in HRTFs can be associated with the various anatomical structures by comparing the wavelength of the sound with the dimensions of the structure (Shaw, 1975b). This means that considerable knowledge about pinna-based spectral features can be gained from looking at the appropriate frequency range in HRTFs.

The pioneer studies by Shaw and Teranishi (1968) and Shaw (1975) have demonstrated that pinna-based spectral features can be found in HRTFs at frequencies above around 4 kHz. Spectral features associated with the torso occur at frequencies below around 2 kHz (Shaw and Teranishi, 1968; Shaw, 1975b; Kuhn, 1987). The main spectral feature associated with the meatus is a non-directional prominent spectral peak at around 2-3 kHz (Shaw and Teranishi, 1968). With regard to spectral features associated with the head, the theoretical results presented above show that the head does not introduce pronounced spectral features, at least under the spherical-shape approximation (see Fig. 2.3). Its main effect is to produce an overall gain on the incident sound. In this thesis, the contribution of each anatomical feature is thoroughly investigated in Chapters 3, 4 and 6. In particular, experimental measurements of the spectral contribution by the head are presented in Chapter 3.

A direct consequence of these results is that much information about specific characteristics of pinna-based spectral features and their dependence on source location can be derived from looking at the high frequency end of HRTFs ($f > 4$ kHz). This is precisely one of the aims of this thesis (Chapters 3 to 6). New HRTF data will be provided with a view to examining pinna-based spectral features in search of common spectral patterns across individuals. The new HRTF data will be also compared with a large collection of HRTF data already available in the literature obtained under various conditions and source locations (for instance, Shaw, 1966; Shaw and Teranishi, 1968; Shaw, 1974; Hebrank and Wright, 1974; Shaw, 1975; Shaw, 1975b; Mehrgardt and

The role of pinna-based spectral cues in sound localisation

Some pinna-based spectral features observed in the frequency transfer function of the external ear (see below) have been found to vary systematically with source location, which suggests that they may signal the location of the sound source. For instance, with regard to vertical localisation, Hebrank and Wright (1974), Butler and Belendiuk (1977), Watkins (1978) and Bloom (1977) have provided strong evidence that narrow elevation-dependent spectral notches are responsible for the sensation of source elevation. Blauert (1969/70), however, has provided evidence that different 'boosted frequency bands' determine the perception of the source in front, overhead or behind.

The role of pinna-based spectral cues is not so well established with regard to sound localisation in the horizontal plane. As explained above, it has been observed that sound localisation in the horizontal plane relies mainly upon head-based ITDs and ILDs (see Middlebrooks and Green, 1991). However, Musicant and Butler (1984) reported that pinna cues have indeed a double contribution to localisation in the horizontal plane: (1) they aid in resolving front/back confusion, and (2) they increase the localisation accuracy when locating sounds within the same quadrant of the horizontal plane.

The problem of developing a mathematical description of the transfer function of the pinna

The directional filtering effect of the pinna is the result of various acoustic phenomena such as reflection, shadowing, dispersion, diffraction, interference and resonance (Blauert, 1983; Kuhn, 1987). The complexity of such a system is increased by its intricate geometry and dimensions.

Because of the system's complexity, very few attempts have been made to model the physical phenomena occurring in the pinna (for a comprehensive review see Blauert, 1983 pp. 63-69), let alone to provide a complete mathematical description of the
pressure field at the meatus entrance. At most, existing models consider the pinna as a sound reflector where only one or two delays are considered between the direct sound wave entering the meatus and the reflected waves from the pinna walls. Batteau (1967, 1968) proposed one such model of the pinna in which the total pressure at the meatus entrance is the sum of the direct, unreflected pressure wave plus two reflected waves, each with a different delay. In Batteau's model, different source locations generate different delays for the reflected waves (see above). Watkins (1978) implemented Batteau's model with the delays measured by Batteau and Plante (1962) for elevation-varying sources. With this approach, Watkins synthesised broad-band stimuli whose spectral shapes were a reasonable approximation to measured HRTFs for sources in the median vertical plane, particularly concerning spectral notches.

Although Watkins' work supports the validity of Batteau's model to some extent, the fact is that such a model is too simplistic (Blauert, 1983). Because the dimensions of the pinna are smaller or comparable to the sound wavelengths, sound diffraction must play an important role in determining the actual pressure field at the meatus entrance. This means that, at the very least, a large number of delays, and not only the two proposed by Batteau, should be properly included in the formulation. In addition to this, Batteau's model does not explain how the dependence of the transfer function of the pinna on source location comes about.

In an attempt to shed some light on the latter issue, Hebrank and Wright (1974) speculated that reflections in the concha (Fig. 2.4) alone may play an important role in generating the elevation-dependent pinna-based spectral features (specifically notches) observed in HRTFs for sources in the median vertical plane. In their view, interference between the direct and the reflected waves from the posterior wall of the concha explain the characteristic elevation-dependent spectral notches. However, their formulation is also very simplistic and does not include diffraction phenomena in the concha. In their model (see Chapter 5), the total pressure at the meatus entrance is simply expressed as the sum of the direct and the reflected (delayed) sounds from the concha posterior wall. The length of the reflected path, hence the delay, varies with the elevation of the sound source, which explains the dependency of the centre frequency of the notch on the elevation angle.

An important objective of this thesis is to investigate how the dependency of pinna-based spectral features on source location is generated in the pinna. Our investigation
Chapter 2: Background

will be done in two ways: (1) theoretically, by producing a physical model of *diffraction, multiple reflections* and *interference* phenomena in the concha (Chapter 5); and (2) experimentally, by measuring the spectral characteristics associated with individual anatomical features of the pinna and their dependency on source position (Chapter 6).

2.3 AUDITORY NERVE ENCODING OF PINNA-BASED SPECTRAL CUES

So far, we have reviewed the evidence that supports the view that the pinna plays an important role in sound localisation because it generates location-dependent spectral cues. We have also reviewed some psychophysical evidence that supports the view that pinna cues are likely to be used by the auditory system in the form of spectral cues. From a purely physiological point of view, this evidence raises a number of interesting questions. In the first place, how (in what form) are pinna-based spectral cues signalled to the central nervous system (CNS)? and, secondly, to what extent and under what conditions is the location information present in pinna-based spectral cues signalled to the CNS? The investigation of these issues constitutes the second major objective of this thesis.

If pinna-based spectral cues are signalled to the CNS in some way, they must be necessarily encoded in the discharge patterns of auditory nerve (AN) fibres. Therefore, the problem may be reduced to investigating possible ways in which pinna-based spectral cues are encoded in the activity of the AN. To date, there have been very few experimental studies addressing this issue directly (Poon and Brugge, 1993a,b; Rice *et al.*, 1995). (These studies will be fully reviewed below). However, there have been numerous studies on the AN coding of speech signals (Sachs and Young, 1979; Sachs, 1984; Delgutte, 1980; Delgutte and Kiang 1984a,b; Sachs *et al.*, 1988). A parallelism may be established between the problem of AN encoding of pinna-based spectral cues and that of AN encoding of speech signals, in the sense that studies of speech perception have almost always relied upon formant frequencies as the basis for the analysis of speech signals (Liberman *et al.*, 1967). Therefore, some of the knowledge about the nature of the AN encoding of pinna-based spectral features can be obtained from reviewing the studies on the AN encoding of speech signals.
For instance, these studies (Sachs and Young, 1979; Sachs, 1984; Delgutte, 1980; Delgutte and Kiang 1984a,b; Sachs et al., 1988) have investigated two possible representations of the spectrum of speech signals in AN discharge patterns: (a) in terms of the profile of the discharge rate of AN fibres as a function of best frequency [known as rate-place representation (Sachs and Young, 1979)], and (b) in terms of the temporal aspects of fibres response due to phase-locking phenomena [known as temporal-place representation, (Young and Sachs, 1979; Sachs, 1984)]. Because phase-locking occurs only for frequencies below around 4 kHz (Palmer and Russell, 1986), it is unlikely that pinna-based spectral features (which occur above around 5 kHz) are encoded in the temporal aspects of AN activity. Therefore, the AN encoding of pinna-based spectral features must be done exclusively in terms of the discharge rate of the fibres (Poon and Brugge, 1993a; Rice et al., 1995).

2.3.1 Rate representations of pinna-based spectral features

Experimental results and considerations

Stationary stimuli

Sachs and Young (1979) studied the AN rate representation of vowel sounds ($f < 5$ kHz). Their most interesting result was a decrease in the quality$^4$ of the AN rate profile with increasing stimulus intensities. Similar results have been reported only recently for pinna-based spectral cues ($f > 4$ kHz) (Rice et al., 1995). Sachs and Young (1979) attributed this result to three possible mechanisms; the broadening of the cochlear filters at high stimulus intensities, the saturation in the discharge rate of AN fibres, and two-tone suppression phenomena, in the sense that energy at the formant peaks may suppress the activity of the fibres with BFs at another peak. The same mechanisms must apply in the case of pinna-based spectral cues, even though they occur at higher frequencies.

$^4$ In this context, the word 'quality' must be understood to refer to the degree of similarity between the stimulus spectrum and the AN rate-profile representation.
Although it is likely that all three mechanisms proposed by Sachs and Young (1979) have a 'negative' effect on the quality of AN rate profiles, it is still unclear to what extent each of them is responsible for the deterioration of the AN rate representation of the pinna-based spectral features at high signal levels. Additionally, the results found by Sachs and Young (1979) for vowel signals, and Rice et al. (1995) for HRTF-filtered stimuli, raise an interesting question: how is the spectral information (if at all) encoded in the AN discharge patterns at high signal levels?

With regard to the first question, it is usually accepted that the dynamic range of AN fibres is the most limiting factor. Consequently, it has been speculated that low spontaneous-rate (LSR) fibres are more appropriate to encode the stimulus spectrum at high stimulus levels, as they have higher thresholds and wider dynamic ranges (Sachs and Young, 1979; Sachs and Abbas, 1974; Winter et al., 1990; Liberman, 1988). Sachs and Young (1979, figure 12) observed that the quality of the rate representation associated with LSR fibres improves slightly with respect to that of high spontaneous-rate (HSR) fibres but deteriorates nevertheless with increasing signal levels. Sachs and Young attributed this deterioration to two-tone suppression phenomena. In a similar line of thought, it has been reported that the AN rate representation of the stimulus spectrum is better at the stimulus onset than in the steady-state (Delgutte and Kiang, 1984a,b). This improvement has been attributed to the wider dynamic range of the onset response.

The issue could be settled by measuring the cochlear response to complex stimuli simultaneously at a large number of points along the basilar membrane (BM) at various signal levels. Unfortunately, no such measurements have been reported because of the limitations of current experimental techniques (for a review see the Introduction of Sellick et al., 1982). An alternative approach is to model this response, as adopted in this thesis (Chapters 7 and 9).

**Moving stimuli**

Another interesting issue is whether or not the AN rate representation of the stimulus spectrum improves when changing between two different conditions. This would be, for instance, the case of a moving source. Poon and Brugge (1993a) studied the response of single AN fibres (in cats) to moving notches like those found in cat's HRTFs (Rice et al., 1992). They found that a moving-notch stimulus, but not a
stationary one, increases the sensitivity of HSR fibres by reducing the discharge rate below the spontaneous levels when the notch is at or in the vicinity of the fibre's BF. This finding led them to conclude that localisation ability may be enhanced when there is relative motion between a sound source and the head.

After the studies by Poon and Brugge (1993a,b), Rice et al. (1995) suggested that "...the nature of the neural representation of a particular stimulus spectrum cannot be inferred from those data (i.e., data corresponding to single AN fibres); such understanding necessarily requires a population study of the responses to one or a few spectra of a representative group of AN fibres." For this reason, Rice et al. (1995) measured the response of a large population of AN fibres to HRTF-filtered broad-band noise. However, Rice et al. did not measure the response to moving stimuli directly. In an attempt to shed some light on the response to moving sources, they presented a related result; they examined the rate difference between the AN rate profiles of two stationary HRTF-filtered stimuli and found a considerable improvement with respect to the individual representation of stationary stimuli.

Unfortunately, because of its experimental design, the study of Rice et al., (1995) does not provide any information about transient AN representations for a large population of fibres when the stimulus changes from one location to another. This transient response will be determined by the degree of adaptation of AN fibres, but from the evidence provided above it seems logical to hypothesise that changes in energy will be enhanced in the transient AN rate representation. This hypothesis is also investigated in this thesis (Chapter 9) by modelling the response of a large number of fibres to HRTF-filtered stimuli and analysing the transient response between different HRTFs.

2.3.2 Modelling the AN rate response to HRTF-filtered stimuli

A number of unresolved issues have been presented above with regard to the AN representation of pinna-based spectral cues. Hypotheses have been formulated which may explain those issues. In most cases these hypotheses could be tested by experimentation. However, currently available techniques are limited and, therefore, there are some instances (see above) in which much progress can be made by modelling the AN response to HRTF-filtered stimuli. The advantages of computer models in auditory research has been discussed by Hewitt (1992). An important aim
of this thesis is to develop and test a full model of the auditory periphery. The model is then used to investigate the AN rate representation of pinna-based spectral cues.

It is important to notice that the controversy about the AN rate representation of pinna-based spectral features arises because of the fact that the cochlear filtering is nonlinear (Rhode, 1971). Cochlear nonlinearities generate two-tone suppression phenomena which, as explained above, may be an important determining factor of the quality of the AN rate representation of the stimulus spectrum. Therefore, if a modelling approach is to be employed to shed some light into the unresolved experimental issues, the model must necessarily include the nonlinear cochlear filtering as an essential characteristic. Furthermore, the model has to simulate two-tone suppression phenomena. Such a model of the cochlear filtering has been provided only recently by O'Mard and Meddis (1996).

An important part of this thesis is dedicated to developing and testing a comprehensive computer model of the auditory periphery which includes the nonlinear auditory filter bank proposed by O'Mard and Meddis (1996). The model is used to investigate the AN rate representation of pinna-based spectral features both for stationary stimuli and moving stimuli, in an attempt to shed some light on some of the obscure issues presented above with regard to the AN rate representation of the pinna-based spectral features. Special attention is given to investigate (a) what is the most limiting factor in the deterioration of the quality of the AN rate representation at high signal levels, (b) possible mechanisms to improve the quality of the representation at high signal levels, and (c) the transient response as the stimulus moves from one HRTF to another.

One important characteristic of our study is that human HRTFs are used for the first time, as opposed to HRTF-like or cat HRTFs which have been used in previous studies (Poon and Brugge, 1993a,b; Rice et al., 1995).

---

5 The HRTFs employed in our simulation were actually measured with a KEMAR, pinna model DB-061 (see Chapters 7 and 9).
CHAPTER 3

Experimental head-related transfer functions measured using a KEMAR fitted with individualised moulded pinnae

3.1 INTRODUCTION

Head-related transfer functions (HRTFs) can be measured either using real people (e.g., Carlile and Pralong, 1994; Pralong and Carlile, 1994; Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989; Middlebrooks et al., 1989) or using a Knowles Electronic Manikin for Acoustic Research — KEMAR — (Burkhard and Sachs, 1975; Gardner and Martin, 1995). The main argument in favour of using real subjects is that the spectral features related to the specific anatomy of each subject (particularly those related to the pinna) are preserved in the results obtained. This important aspect is not guaranteed using a KEMAR. However, an attempt to obtain approximated individualised HRTFs using a KEMAR can be made simply by using individualised moulded pinnae fitted on the manikin's head instead of the standard pinnae provided by the KEMAR manufacturers.

The use of a KEMAR for making HRTF measurements has some added advantages over using real subjects (Burkhard and Sachs, 1975). Some of these advantages are: (1) the possibility of exchanging the pinnae easily, (2) the possibility of producing alterations (e.g., removal of certain anatomical structures) of the pinna shape, (3) the possibility of using pinna-replacements for measuring, for instance, the HRTF associated with the head only, without including the effect of the pinna, (4) the reproducibility of the results, (5) the fact that subjects are not required to spend long recording sessions.
A method for recording HRTFs using a KEMAR (model DB-4004) fitted with individualised pinnae is presented in this chapter. The experimental method for moulding the pinnae is explained in detail. Experimental results obtained with this technique are presented. The transfer function of the head together with an ear-canal simulator (Knowles simulator, model DB-100) is presented. The experimental transfer function of the pinna alone is also shown. A comparison of the HRTFs obtained for several moulded pinnae is made and the results discussed.

3.2 EXPERIMENTAL TECHNIQUE FOR MOULDING REAL PINNAE

A substance called Otoform is used to mould the pinnae. (Otoform is normally employed for taking impressions of hearing aids.) Otoform is a soft and "sticky" silicon-based substance which does not harden unless mixed with an appropriate amount of hardener (see below).

The process for producing a full pinna model has three stages:

1. Making a "negative" pinna mould by pressing Otoform around the subject's ear-canal, pinna and surrounding area.
2. Making a pinna impression ("positive" model) by injecting Otoform into the negative pinna mould obtained in stage 1.
3. Fixing the pinna model produced in stage 2 to the KEMAR mould. The KEMAR mould is a cast of one of the depressions at both sides of the manikin head where the pinnae are fitted into.

The final product in stage 3 is a full pinna model which fits into the KEMAR's pinna placement. Only one pinna is moulded at a time.

---

1 Otoform-K2, 800g, Condensation-Vulcanising Silicone Impression Material with Hardener, Cat. No. 071K2 is commercialised by P. C. Werth Ltd., 45 Nightingale Lane, London, SW12 8SP, Fax: 081 675 7577.
3.2.1 Health and safety measures

The subject is instructed to read and sign (if appropriate) a consent form so that the pinna moulding process can be carried out. In the interest of hygiene hands are thoroughly washed before and after model making. The condition of the subject's ear canal and tympanic membrane is examined using an Otoscope before proceeding to making the negative mould of the pinna. A fresh sterilised Otoscope nozzle is used for each ear-canal inspection. If any sign of inflammation, redness or discomfort is observed in the subject the procedure is halted.

3.2.2 Making the 'negative' pinna mould

The moulding compound is prepared by thoroughly mixing 80 g of Otoform with 0.5 ml of hardener. This quantity of Otoform is enough for one pinna mould only. The more hardener is added to the 80 g of Otoform, the quicker the compound dries up and the harder the final mould becomes. The compound is then squeezed into a 100 ml syringe. The subject is asked to place an oto-stop in his/her ear-canal as far down as is comfortable. The oto-stop prevents the Otoform compound from damaging the eardrum. With the subject's head on the side, the moulding substance is gently squeezed into the subject's ear-canal, working outwards and gradually filling the ear-canal, concha and pinna-flange. When the pinna is entirely covered with a substantial amount of compound, the subject is asked to remain in this position for approximately 20 min. After this time the mould can be removed from the subject with care (provided it is sufficiently dry) and left to reach full rigidity.

3.2.3 Making the pinna impression

Once the negative pinna mould has hardened sufficiently (approximately 48 hours), the positive pinna impression can be made. In this case the moulding compound is obtained by mixing 30 g of Otoform and 0.15 ml of hardener. The mixture is squeezed into a 10 ml syringe. The negative mould should be covered with a thin layer of Vaseline to prevent the new Otoform mixture from sticking to the negative mould (notice they are made of the same substance). The Vaseline can be applied using a cotton bud. The moulding compound can be now injected into the negative
pinna mould using the syringe. It is important to avoid the formation of air bubbles at this stage. The part which will eventually be the ear canal must be also covered. After approximately 18 hours the pinna impression can be extracted from the negative mould taking care not to tear it at the more intricate parts. The pinna impression should be cleaned and perfected (if necessary). Any imperfection, such as small holes, can be smoothed to the contours of the pinna by filling them up with more Otoform compound.

3.2.4 Fixing the pinna impression to the KEMAR mould

The pinna impression obtained in the previous stage can not be fitted to the KEMAR yet. It must first be fixed to a KEMAR-fit. A KEMAR-fit is a model impression which fits perfectly in either one of the two square recesses on both sides of the manikin head where the pinnae are usually fitted into.

The KEMAR-fit is an impression obtained from a KEMAR-mould. The KEMAR-mould was produced as follows. A cast of one of the square recesses was obtained directly from the KEMAR using Otoform/hardener compound (80 g of Otoform, 0.5 ml of hardener). The ear-canal simulator (Knowles occluded-ear simulator, model DB-100) was in-place when making the cast. The KEMAR-mould was then produced from this cast using a substance much harder than Otoform (e.g., Isopon Car Body Filler). The KEMAR-mould is, therefore, a faithful impression of the depressions at either side of the manikin head where the pinnae are placed. The KEMAR-mould can be reused each time a full pinna model is to be produced.

The advantage of having two independent casts, namely, the pinna impression ("positive" cast) and the KEMAR-mould, is that the pinna impression can be accurately fixed to the KEMAR-fit at the specific angle of the subject's pinna. In fact, the KEMAR-mould is already set back slightly at an angle to facilitate the process. Two KEMAR-moulds, one for each ear, are needed. They obviously differ in the slight angle at which they are set back.

The process of fixing the pinna impression to the KEMAR-fit is as follows. A neat cut is made around the pinna impression to remove the excess material. Sufficient Otoform compound (30 g of Otoform, 0.15 ml of hardener) is mixed and pushed into
the KEMAR-mould. The pinna impression is then laid on the top at the correct angle. The pinna impression is then pushed down into the mixture causing it to spill over the edge of the KEMAR-mould. The fingers (or a spatula) are used to smooth the Otoform compound to match the pinna impression taking off any excess material, if necessary. The product is left to dry up for about 48 hours. A full pinna model which fits into the appropriate (left or right) square recess of the KEMAR is finally obtained. It only remains to perforate the ear-canal entrance from the full pinna model. This is done with a circular chisel with a diameter of 7.5 mm to match the diameter of the ear canal simulator.

3.2.5 Pinna replacements: the 'infills'

*Quasi-flat* pinna-replacements were produced from the KEMAR-mould only. These replacements will be referred to as the infills (see Appendix 1). They are slightly funnel-like to gradually smooth the sound pathway from the head surface to the actual meatus entrance (i.e., the entrance of the earcanal simulator). When replacing the manikin pinnae, they produce the effect of having no pinnae but rather having a pair of holes at both ear-canal entrances. They are used to measure the head-shadow effect, for instance (see below).

3.3 EXPERIMENTAL METHOD AND EQUIPMENT FOR MEASURING HRTFs

All head-related transfer functions presented in this chapter were measured using a KEMAR (model DB-4004) fitted with Knowles occluded-ear simulators (model DB-100) and 1/2" condenser Bruel & Kjaer microphones (model 4132). This equipment is specially designed to match the average acoustic response of a human listener at the eardrum (Burkhard and Sachs, 1975). Moulded pinna auricles or pinna replacements (infills) were used depending on the task of the experiment. Unless otherwise stated, the measurements were made outdoors under quasi free-field conditions (see below).
The experimental set-up and equipment are shown in Fig. 3.1. The KEMAR was mounted onto a 1 degree-scaled rotating vernier base which allows azimuthal variations in the position of the KEMAR relative to the speaker from 0° to 360°. The speaker is mounted onto a speaker-stand whose height can be varied to achieve the desired elevation. Unless otherwise stated, the distance between the speaker and the KEMAR's head was 1.4 m. This is the minimum distance required to assume confidently that the wave reaching the manikin is a plane wave.

Figure 3.1. Experimental set-up and equipment for HRTF measurements.
3.3.1 Stimulus generation: the super-pulse

The delivery/recording system employed to make HRTF measurements does not have a flat frequency response. Consequently, the frequency response of the delivery/recording system needs to be compensated for in order to measure the actual head-related transfer function corresponding to a particular source location. Computer code was available\(^1\) which allowed us to record the impulse response (defined as the system's response to an ideal click) of the delivery/recording system using a Masscomp computer (model MC5450). [It must be pointed out, however, that although the HRTF measurements presented in this Chapter were obtained with the existing system, i.e., with the Masscomp computer and the available code, an improved method was developed during the course of this research. New computer code was written and the whole process was reconstructed and thoroughly checked for the system to run using the audio facilities of Silicon Graphics (Indy) computers. The improved method is described in Chapter 7]

With the original system, the impulse response of the system was measured with one of the two B&K microphones held by a standing clamp positioned in quasi free-field conditions. It was assumed that the frequency response of both B&K microphones was identical (as claimed by the manufacturer). The same computer code creates a click-like stimulus (referred to as "the super-pulse") whose spectrum is complementary to that of the impulse response of the delivery/recording system. The idea being that the spectrum of the system's response to the super-pulse stimulus must be flat. Accurate HRTF data can be obtained simply by using the super-pulse as the stimulus.

The way in which the super-pulse was generated can be summarised as follows:

1. A specified number of Dirac clicks (at least 40) are generated by the software. The clicks are transmitted to the speaker using the built-in DAC of the Masscomp. The clicks are delivered by the speaker having previously been amplified using the analogue amplifier (see Fig. 3.1). The impulse response of the delivery/recording system is recorded (for each click) by one of the B&K microphones, amplified and input to the Masscomp through the ADC.

---

\(^1\) The code was written by Trevor Shackleton while he was working at the Speech and Hearing Lab., Dept. of Human Sciences, (Loughborough University).
2. The average impulse response of the system is calculated in the time domain (see Fig. 3.2a). The idea behind the averaging process is that the ambient random noise present in quasi free-field conditions will cancel if the number of clicks is sufficiently large. The actual number of clicks varied from one experiment to the next, depending on the amount of ambient noise (mainly caused by the wind). In most cases 40 clicks were used.

3. An FFT of the average signal was calculated (see Fig. 3.2b). This represents, of course, the average frequency response of the delivery/recording system. This spectrum would be flat for an ideal system.

4. A mirror image of the average frequency response is calculated (see Fig. 3.2c).

5. An IFFT of the mirror-image spectrum is calculated (see Fig. 3.2d). The resulting time-domain signal is named Super-pulse.

6. Steps 1 to 3 are repeated using the super-pulse as the stimulus instead of an ideal click. Fig. 3.2e shows the time-domain average response of the system to the super-pulse.

7. System check. An FFT of average response obtained in step 6 is calculated in order to check the spectral variations of the system’s response to the super-pulse. The spectrum of the average response is now approximately flat with variations of ±2 dB in the specified range of frequency (see Fig. 3.2f).

To achieve a super-pulse which generates a flat-spectrum response for the delivery/recording system is almost impossible in practice (see Fig. 3.2f). Unless otherwise stated, spectral variations of ±2 dB within the frequency range of interest (see below) were accepted for most experiments. Since most experiments investigate well-marked spectral features with amplitudes of the order of ±15 dB or above, these small variations do not lead to a misinterpretation of the data. This issue is discussed further in the discussion section below.
Figure 3.2. Super-pulse generation technique. (a) Time-domain average response of system to a specified number of Dirac pulses. (b) Spectrum of the average time response shown in (a). (c) Spectral mirror image of (b). (d) The super-pulse: time-domain signal corresponding to the mirror spectrum shown in (c). (e) Average system's response to N super-pulses. (f) Spectrum of the average system's response shown in (e). Note that the frequency range is from 1 to 10 kHz in this case. Note the spectrum of the average response of the system to N super-pulses (f) has spectral variations of ±1 dB in this case.
Other measures adopted to improve the quality of the super-pulse

The following measures were taken in order to improve the quality of the super-pulse:

a) Echoes were eliminated by setting up the experiment outdoors under quasi free-field conditions. Both the speaker and the microphone were placed onto a low-reflecting grass surface. Additionally, the recording time was set short enough (20 ms) to prevent echoes from being recorded.

b) Experiments were never performed on windy days, since wind introduces an enormous amount of random noise difficult to compensate for.

c) A number of stimuli (no fewer than 40) were used in an attempt to cancel out random ambient noise. This, therefore, helped to increase the signal-to-noise ratio in the average signal.

d) A more obvious way to increase the signal-to-noise ratio was by increasing the signal intensity using the volume control in the amplifier. However, this was possible only up to a certain extent. Excessively high volumes produced an overload (clipping) in the input port of the Masscomp computer. To solve this problem a pre-attenuator (introducing either 10 or 20 dB attenuation) was connected between the microphones and the input to the analogue filter (see Fig. 3.1). Consequently, the use of the pre-attenuator increased the signal-to-noise ratio by 20 dB and also avoided the clipping problem.

e) Even though the software should compensate for the frequency response of the speaker, it was found that different speakers were more suitable to generate a good super-pulse at different elevation angles. The reason why this happens is unclear but a possible explanation may be given in terms of the directional characteristics of the speakers. Two different speakers were used, namely a KEF (C-35) speaker and a Radio Spares (RS) Wide-Range Twin-Cone 6" speaker. The KEF speaker was most suitable for low elevation angles (−40°) having a frequency response relatively flat between 1 and 17 kHz. For higher elevations (0°), however, its frequency response shows a deep spectral notch of around −40 dB at 10 kHz, for which the super-pulse generation software could not compensate. The RS speaker was found to generate a better super-pulse at this elevation (0°).
3.3.2 HRTF measurements

Actual HRTFs measurements were made only after a super-pulse had been generated which met the ±2 dB specification. The measurements were produced using the set-up and equipment described above (see Fig. 3.1) with the KEMAR wearing either individualised moulded pinnae or the infills. The ear-canal simulators were used in all experiments. All experiments were done with the manikin wearing neither clothes nor a wig.

HRTF measurements were made following steps 1 to 3 of those explained in the super-pulse generation procedure (see section Stimulus Generation: the Super-pulse above). However, the super-pulse was used as the stimulus this time, instead of the ideal click employed when generating the super-pulse. In other words, for each position of the sound source, the resultant head-related transfer function was presented as the spectrum of the average time-domain impulse response obtained using the super-pulse as the stimulus. It is important to notice that a new super-pulse was generated at the beginning of each experiment, whenever there was a change in the experimental conditions (such as the distance between the speaker and the KEMAR) or whenever a piece of the delivery/recording equipment was changed during the recording session (for instance, different speakers were used at different elevation angles). Each new super-pulse was only accepted if it generated a quasi-flat system response, with variations smaller than ±2 dB.

The frequency range of the HRTF spectral representations differed depending on the experiment. In the early experiments the frequency range was set to 1-10 kHz. In later experiments it was expanded to 1-17 kHz, in order to observe the spectral transformations introduced by the pinna and, particularly, by the concha. HRTF data were obtained with a frequency step $\Delta f = 31.73828$ Hz. However, only one out of 4 samples (thus $\Delta f = 126.95312$ Hz) was represented in the results figures (below) due to limitations in the graphics computing power.\(^3\)

---

\(^3\) The graphical representation of the results (see below) was done using the computer program Deltagraph 1.5 (UK) which only runs on 16-bit Macintosh computers.
3.4 SYSTEM VALIDATION: RESULTS

In this section, HRTF data obtained with the method explained above are presented. Firstly, results obtained with the KEMAR fitted with individualised moulded pinnae are shown. Results for several pinnae shapes are presented and compared when possible. Secondly, HRTF data obtained without the pinna are shown. Thirdly, and lastly, the approximated transfer function of a moulded pinna is presented. This approximated transfer function is calculated by subtracting the results for the no-pinna condition from those obtained with KEMAR fitted with pinnae.

In this section, HRTF measurements were made only for azimuth-varying sources. Measurements were made every 5°, for azimuth angles between 270° and 90° in the frontal hemisphere (see section Angle conventions in Chapter 1). Unless otherwise stated, the speaker was positioned at an elevation angle of -40° and the distance between the speaker and the KEMAR was 1.4 m.

No measurements were made for elevation-varying sources at this stage for two reasons. Firstly, the experimental set-up did not allow the variation of the elevation angle in a systematic and accurate way. Secondly, the main purpose of the measurements was to validate the method by comparing our results with HRTF data previously published in the literature, which were mainly presented in the form of azimuth-varying HRTFs (Shaw and Teranishi, 1968; Teranishi and Shaw, 1968; Mehrgardt and Mellert, 1977). (HRTF measurements at different elevation angles are presented in Chapters 4 and 6.)

Because the super-pulse technique was used to collect the data, we are certain that the spectrum of the system's response to the super-pulse is almost flat, with variations of less than ±2 dB. However, in these early experiments we failed to measure the exact pressure level (reference level) at the microphone placed in free field at the usual position of the KEMAR. Different pressure levels may have been used in different experiments, all of them being around 52 dB. Therefore, the figures shown below represent the total pressure at the eardrum instead of the actual transfer function. The transfer function would be obtained by subtracting the pressure level at the standing
microphone (expressed in dB) from the total pressure levels plotted. (Actual transfer functions are shown in Chapters 4 and 6.)

3.4.1 HRTF measurements with the pinnae in place

Measurements were made with moulded pinnae from three subjects (IX, AC and LPO). Figs. 3.3a and 3.3b show the HRTFs obtained with impressions of the left and right pinnae respectively of subject IX. Similarly, Figs. 3.4a and 3.4b show the HRTF results obtained with impressions of the left and right pinnae respectively of subject AC. Finally, results for subject LPO are shown in Figs. 3.5a (left ear) and 3.5b (right ear). The shape and dimensions of both left and right pinnae of subjects IX, AC and LPO are given in Appendix 1. For each source location and subject, measurements were made for the left and the right pinna simultaneously. Notice that the frequency range in the results corresponding to subjects IX and AC only goes from 1 to 10 kHz, whereas it goes from 1 to 17 kHz in the case of LPO.

Results: HRTF measurements with the pinna in place

A comparison of the HRTF results obtained across subjects reveals spectral characteristics which are specific for each subject, particularly at frequencies above 7 kHz (Shaw and Teranishi, 1968). However, a careful analysis show that there are some spectral features common to all of them. An attempt has been made to label the most salient spectral features, using the same label for corresponding spectral features across subjects.

The spectral features common to all pinnae are the following:

- There is a shadowing effect which is more intense at high frequencies (>3 kHz). In other words, high frequencies are largely attenuated when the sound source is in the contralateral side.
- There is a clear spectral peak, P1, centred at around 2.5 kHz. P1 appears for all azimuthal positions of the source, although it is not so intense when the source is in the contralateral side.
- There is a second spectral peak, P2, centred between 3.5 and 5.5 kHz. This peak is very wide and is only present for sounds incident from the ipsilateral side.
• A clear spectral notch (N1) appears at around 7 kHz. For most pinnae, the notch hardly depends on the azimuthal position of the source. In some cases, it appears at slightly lower frequencies for azimuthal angles away from the lateral vertical plane.

• There is another peak, P3, between 8 and 10 kHz. Like P2, P3 only appears for sound sources in the ipsilateral side.

• Despite the fact that no data above 10 kHz was obtained for subjects IX and AC, there is some indication in the available data that the spectral notch N2 (at around 10 to 12 kHz) observed for both ears of subject LPO is also present for both ears of subjects IX and AC. It can be seen in Fig. 3.5 that the centre frequency of this notch depends on the azimuthal angle of the source, going from approximately 10 kHz for a source in the median plane (0°) to 12 kHz for a source in the lateral vertical plane (i.e., 270° azimuth for the left ear and 90° azimuth for the right ear, Fig. 3.5).

Data obtained with the left and right pinnae of subject LPO (Figs. 3.5a and 3.5b) show spectral features above 10 kHz. Another spectral peak, P4, is observed at 14 kHz for a sound source in the lateral vertical plane. Finally, a spectral notch, N3, is observed at around 16-17 kHz. It is worth noticing how the amplitude and centre frequency of P4 and N3 are different for the right and left ears and yet one is able to see a common pattern between them.
Figure 3.3. HRTFs measured with the KEMAR fitted with the pinnae of subject IX as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Notice that the total pressure at the microphone (eardrum simulator) is plotted instead of the actual transfer function (see text for details). Notice that the main spectral features are marked: P = peak, N = notch. N.B., the frequency range goes from 1 to 10 kHz.
Figure 3.4. HRTFs measured with the KEMAR fitted with the pinnae of subject AC as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Legend as in Fig. 3.3.
Figure 3.5. HRTFs measured with the KEMAR fitted with the pinnae of subject LPO as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Legend as in Fig. 3.3. N.B. the frequency range goes from 1 to 17 kHz.
3.4.2 The transfer function of the KEMAR head and torso: HRTFs measurements without the pinnae

An experiment was carried out to investigate how much of the spectral information contained in the experimental HRTFs described above is due to the torso, head and ear-canal. Experimental data were collected using the super-pulse technique. Apart from the actual pressure level delivered to the KEMAR, the experimental conditions and equipment were similar to those present when collecting HRTF data using LPO's pinnae, so that the results in both cases could be directly compared (Figs. 3.5 and 3.6). In this case, however, the KEMAR was fitted with the infills (see section Pinna replacements: the 'infills' above, and Appendix 1) instead of moulded pinnae. HRTF data were collected for the same source positions as above. For each source location, data were collected for the right and left infills simultaneously. Results are shown in Fig. 3.6.

Results: HRTFs measurements without the pinnae

The shadowing effect at frequencies higher than 3 kHz is also present, with the same intensity, in the results obtained with the infill (Fig. 3.6). Therefore, it must be caused by an anatomical part other than the pinna. As a matter of fact, it is quite obvious that the observed shadowing effect must be experimental evidence of the so called head-shadow effect (see Chapter 2: Background).

The spectral peak P1 is also observed in the results obtained with the infills. It appears at the same centre frequency (2.5 kHz) with a similar intensity and source-location dependency as in the HRTF obtained with the moulded pinnae (e.g., see Figs. 3.3, 3.4, 3.5).

The second peak, P2, does not appear so clearly in the HRTFs obtained with the infills, although for azimuth angles between 300° and 345° there is still a noticeable increase in intensity at frequencies comparable to the centre frequency of P2.

The results for the infills reveal a peak at approximately the same frequency as P3 (8-10 kHz). This result indicates that an important contribution to the amplitude of P3 may be generated in the ear-canal.
The spectral notches N1 and N2 clearly observed in the HRTF results obtained with moulded pinnae (Fig. 3.5) are not observed in the results obtained with the infill (Fig. 3.6). There is a drop in intensity around the centre frequency of N1 but no deep minima can be seen. It should be noticed, however, that very local spectral notches centred at frequencies comparable to the centre frequency of N1 and N2 appear in the infills HRTF data. These local notches only appear at 330° for the left infill and 30° for the right infill (notice the symmetry with respect to the median plane). Although there may be a relationship between these notches and N1 and N2, it is important to notice that N1 and N2 not only appear at 325° but rather at all azimuths, even for sounds in the contralateral side (see below).

Above 12 kHz there are remarkable differences between the sets of results obtained for the left and right infills (Figs. 6a and 6b). For instance, there is a peak at 14 kHz for the left infill at azimuthal angles between 270° and 320°. This peak would contribute to the amplitude of P4 measured with the left pinna. However, a similar peak is not so clearly observed with the right infill, even though P4 is clearly present in the results obtained with the right pinna (Fig. 3.5). Additionally, there is a notch at around 16 kHz at all azimuths for both the left and right infills. The depth and azimuth dependency of this notch are generally very different for each infill, although a common feature to both infills is the fact that the notch depth increases symmetrically at 30° with respect to the 0° line. It could be thought that this notch is related to N3. However, it should be noticed that N3 is much deeper.
Figure 3.6. HRTF for the KEMAR fitted with the infills (i.e., no pinnae) as a function of azimuth at -40° elevation (source distance = 1.4 m). (a) Left ear. (b) Right ear. Notice the total pressure at the microphone (eardrum simulator) is plotted instead of the actual transfer function (see text for details). Notice that the main spectral features are marked: P = peak, N = notch. N.B. the frequency range is from 1 to 17 kHz.
3.4.3 An approximated experimental pinna transfer-function

Approximated transfer functions for the left and right pinnae of subject LPO were calculated by subtracting the infill data from the HRTF data obtained with the pinnae in place (data expressed in dB). The results are shown in Fig. 3.7. It is important to notice that the actual pressure levels at the free-field microphone in the two experiments were likely to differ by a few dB. This would imply that the 0-dB level shown in the chart is not absolute. Nonetheless, the results are useful since the relative amplitudes among spectral features are correct.

In Fig. 3.7 it can be seen that P1 is not present in the transfer function of either pinnae. According to Shaw and Teranishi (1968) and Burkhard and Sachs (1975), P1 must be the fundamental (or \( \lambda/4 \), \( \lambda = \) wavelength) mode of resonance of the ear-canal, related to the canal length. This would explain why it is present in the results obtained with and without the pinna (Figs. 3.5 and 3.6 respectively) and does not appear in the pinna transfer function (Fig. 3.7).

From the study by Shaw and Teranishi (1968) it may be concluded that P2 must correspond to the \( \lambda/4 \) "depth" resonance mode of the concha, properly corrected for an open-end cavity. This agrees with the fact that P2 is not clearly observed in the HRTFs measured with the infill, but is present in the data obtained with the pinnae. However, it rises two questions: (1) why is P2 not present more clearly in the approximated transfer function of the pinna (Fig. 3.7); and (2) what causes the increase in intensity at around 4-5 kHz between 300° and 345° for the left infill, and 15° and 45° for the right infill (Fig. 3.6). The answer to the first question is that since our measurements were made at -40° elevation we should not expect an amplitude of more than 6 or 7 dB for a sound source in the lateral plane [see figure 12 in Shaw and Teranishi (1968)]. This fact together with the possibly erroneous 0-dB reference of Fig. 3.7 may have produced a spectral peak with an amplitude between 0 and 5 dB, which in our graphic representation has been plotted as a green patch at 4-5 kHz in Fig. 3.7. Regarding the second question, it may be that the funnel-like structure of the infill acts like a "cavity", thus, generating spectral features (e.g., resonances) at certain

---

\[4\] As both data sets were expressed in dB, the transfer function corresponds to the subtraction. If the data were not expressed in dB, the transfer function would be calculated as the ratio between the data sets.
angles of incidence comparable to those produced by the concha. This argument is also applicable to P3 and P4.

Results by Shaw and Teranishi (1968) suggest that P3 and P4 must be related to transverse resonance modes in the concha. It is interesting, however, that P3 and P4 are clearly visible in the transfer function of the right pinna (Fig. 3.7b) but not in the left one (Fig. 3.7a), particularly P4. It may be that the shapes of LPO’s pinnae, the infills or both are not perfectly symmetrical and therefore generate non-symmetrical HRTFs particularly at frequencies above 6 kHz (Searle et al, 1975).

Regarding spectral notches, it is obvious that N1, N2 and N3 are present in the approximated transfer function of both pinnae (Fig. 3.7) and therefore must be generated by the pinna. N1 shows very little dependence on the azimuthal angle (its centre frequency is approximately 7.5 kHz for sources in the lateral vertical plane and 6.5 kHz in the median plane). The characteristics of N1 coincide with those of a spectral notch described by Shaw and Teranishi (1968). Even its amplitude (approximately 10 to 15 dB) is very similar in both cases. Shaw and Teranishi concluded that the notch is generated in the pinna (possibly in the concha), as it also appeared in their blocked-meatus measurements. Our results agree with their conclusion. As for N2 and N3, no specific mention of them has been found in the literature, but our results show they are important notches generated in the pinna. N2 and N3 are slightly more dependent on azimuth (particularly for the right ear) than N1.
Figure 3.7. Approximated transfer function of the pinna calculated by subtracting the infill HRTF (Fig. 3.6) from the HRTF data for subject LPO (Fig. 3.5). (a) Left pinna. (b) Right pinna.
3.5 DISCUSSION

A method for measuring head-related transfer functions using a KEMAR fitted with individualised moulded pinnae or pinna replacements has been presented. The method has been used to measure HRTFs corresponding to the left and right pinnae of three subjects. Additionally, the transfer function of the KEMAR fitted with quasi-flat pinna replacements has been measured. Approximated transfer functions for a subject's pinnae have been obtained by subtracting the infill data from the pinna data.

3.5.1 Experimental method

The validity of the experimental method above described is supported by the fact that most of the spectral features described above common to all subjects (P1, P2, N1, P3 and P4) are in reasonable agreement with results previously published in the literature by other authors (e.g., Shaw and Teranishi, 1968; Teranishi and Shaw, 1968; Burkhard and Sachs, 1975). However, the actual pressure level delivered to the KEMAR should have been measured for each experiment in order to be able to make a faithful one-to-one comparison between our results and those previously published. This problem is solved in the next chapters (4 and 6), where actual transfer functions are presented.

The discrepancies in the results for the right and left infills above 12 kHz may be caused by differences in the actual shape of the infills. Although every care was taken to make the infills identical (Appendix 1), the left infill appears to be slightly more funnel-like than the right infill, which is flatter. This may affect the results in a way which remains unclear. The same comment applies to the discrepancies between the left and the right pinnae. It is known that there are slight but significant anatomical differences between the left and right pinnae (Searle et al., 1975). These differences are included in the modelled pinnae and may generate the discrepancies between the left and right HRTF at high frequencies. For instance, the left and right pinnae of subject LPO shown in Appendix 1 are slightly different in the size of the concha aperture at the point indicated by the star.
3.5.2 Spectral features

The physical processes which generate the spectral features common to all subjects are known only in part. As explained above, spectral peaks have already been explained in terms of acoustic resonances either in the ear canal (as in the case of P1), or in the concha (as in the case of P2, P3 and P4), (Shaw and Teranishi, 1968; Teranishi and Shaw, 1968; Shaw, 1975). The physical origin of the spectral notches is, on the contrary, yet to be found. Besides validating our experimental method, one of the aims of the experiments presented above was to investigate spectral notches present in experimental HRTFs in a manner different from that of Shaw and Teranishi.

Our results have revealed three clear spectral notches N1, N2 and N3 generated in the pinna. N1 has been described by Shaw and Teranishi (1968) as azimuth-independent (for a source at 0° elevation) but elevation-dependent (for a source in the lateral vertical plane). The first observation has been confirmed by our results for all six pinnae employed, although a more careful analysis reveals that there is indeed some slight dependence on azimuth. The elevation dependency of the notch has not been tested here (in this chapter), but has been repeatedly confirmed by other authors for sources in the median vertical plane (Hebrank and Wright, 1974; Butler and Belendiuk, 1977) and in the lateral vertical plane (Bloom, 1977). All these authors have reported the existence of a spectral notch whose centre frequency increases from approximately 6.5 kHz at −45° elevation to 10 kHz at 45° elevation.

Despite all the information available on N1, its physical cause remains to be found. Shaw and Teranishi (1968) speculated that it may be the result of an interference effect between the two successive modes of resonance in the concha, P2 and P3. A second possible explanation was proposed by Hebrank and Wright (1974). They speculated that N1 may be the result of destructive interference between the direct wave entering the meatus and the reflected wave from the posterior wall of the concha. Neither hypothesis have been formally tested, either experimentally or mathematically, but as explained by Hebrank and Wright, the concha dimensions support the concha reflections hypothesis, at least for sounds incident from the median vertical plane.

No discussion about N2 or N3 has been found in the literature, perhaps because most of the previous studies were confined to a smaller frequency range or to different
source locations for which N2 and N3 are not observed. Our results reveal that N2 and N3 show an azimuth dependency similar to that of N1. No experiments have been carried out to measure the elevation dependency of these notches. If the elevation dependency of N2 and N3 was also similar to that of N1, it would be likely that N2 and N3 correspond to higher-order cancellation frequencies of the fundamental N1. In other words, if $f_{N1}$ is the frequency of N1, then N2 and N3 would correspond to $3f_{N1}$ and $5f_{N1}$ respectively. The elevation dependency of N1, N2 and N3 is studied in the next chapter.
CHAPTER 4

Experimental head-related transfer functions measured using a KEMAR fitted with standard pinnae

4.1 INTRODUCTION

This chapter is mainly concerned with investigating the elevation dependency of the spectral features (particularly the notches) described in Chapter 3. Unfortunately, the experimental technique described in Chapter 3 was no longer available to us due to a fatal failure of the Masscomp computer employed to make the head-related transfer function (HRTF) measurements. For this reason, elevation measurements for the same moulded pinnae employed in Chapter 3 could not be made. Instead, we decided to use the head-related impulse responses database provided by Gardner and Martin (1994, 1995). This database contains head-related impulse responses of a KEMAR fitted with standard Knowles pinnae (left pinna model: DB-061; right pinna model: DB-065, see Appendix 1) at 710 positions of the sound source around the head. The data were released in the form of raw head-related impulse responses (HRIRs) and, therefore, a complete frequency analysis was required to express the data as HRTFs. The experimental method employed by Gardner and Martin to record the HRIRs is not explained here for two reasons: firstly, because it has been published elsewhere (Gardner and Martin, 1995) and, secondly, because it is not an original piece of work by the author of this thesis. The frequency analysis of the provided HRIRs and the

1 A fatal failure of the Masscomp computer prevented us from using the technique described in Chapter 3 to investigate the elevation dependency of the spectral features—which is mainly what this chapter is concerned with. However, towards the end of the research process, new code was written so that a similar, but improved, technique for measuring HRTF could be run with Silicon Graphics (Indy) computers. This technique is described and tested in Chapter 6. Results with the new technique at different elevation angles are also presented in Chapter 6.
interpretation of the resulting HRTFs are, on the contrary, original work by the author of this thesis and are, therefore, explained below.

The resulting HRTF data are used in two ways:

1. To compare the spectral features obtained with standard pinnae with those described in Chapter 3, i.e., obtained with individualised moulded pinnae.
2. To examine the elevation dependency of the HRTFs, particularly the spectral notches described in Chapter 3.

4.2 ANALYSIS METHOD FOR OBTAINING THE HRTFs

The frequency analysis of the HRIRs provided by Gardner and Martin (1994, 1995) was performed using the fast Fourier-transform (FFT) function included in the computer package SoundEdit Pro (version 1.0.5). The HRIR data files needed to be converted from their original binary format to AIFF format so that they could be read into SoundEdit Pro. The conversion was done for each file individually by a short C program which makes use of the file facilities of LUTEar (LUTEar, 1993). The parameters required for the conversion were set according to the specifications given by Gardner and Martin (1994, 1995) (i.e., the number of samples was set to 512, the sampling rate was set to 44.1 kHz and the data were read as 16-bit signed integers).

The spectrum of each HRIR was calculated by using the Spectrum menu option (i.e., the FFT function) available in SoundEdit Pro. Unless otherwise stated, the following Spectrum Options were used: Frequency range: 0 to 17 kHz; Gain: 0 dB; Window: Hamming; Offset: 16 points; FFT-size: 256 points (which yields a spectrogram with frequency steps Δf=172 Hz). For each HRIR a full spectrogram was exported as a tab-delimited text file for further analysis with Microsoft Excel. The spectrum, A(f), of each HRIR was obtained from the spectrogram at the time-instant of maximal spectral energy. The spectrum, A(f), was then expressed in dB as follows: \( A_{dB}(f) = 20 \cdot \log_{10}(A(f)) \). The spectrum of the impulse response of the Optimus Pro 7 speaker (the source used by Gardner and Martin), S(f), was calculated and expressed in dB, \( S_{dB} \), in exactly the same manner.
The head-related transfer function for each source location was then obtained by subtracting the speaker contribution (expressed in dB) from the spectrum (also expressed in dB) of the HRIR corresponding to that location. This operation also neutralises the speaker contribution which is included in the original HRIRs. (Notice that the contribution of the Realistic Optimus Pro 7 speaker employed by Gardner and Martin to make their recordings was not compensated for in the original HRIRs). Additionally, it should be noticed that Gardner and Martin used different microphones for measuring the HRIRs (Etymotic ER-11) and the impulse response of the speaker (Neumann KM-84i). The response difference between the ER-11 and the KM-84i microphones has not been neutralised in the resultant HRTFs, but is almost negligible (±2 dB) in the frequency range from 0.5 to 12 kHz. Between 12 and 17 kHz, however, the response of the ER-11 microphone tails down around 6 dB, whilst the KM-84i remains almost flat (the manufacturers specifications for both microphones are shown in Appendix 4). \( S(f) \) was not found to be zero for any frequency within the studied range. On the contrary, \( A(f) \) was sometimes zero for a certain frequency, \( f_0 \). In that case, it was understood that a deep minimum should appear in the HRTF at \( f_0 \) and, therefore, the value of the HRTF was set to \(-100 \text{ dB}\) at that particular frequency.

4.3 RESULTS

4.3.1 A comparison between the HRTFs obtained with individualised moulded pinnae and those obtained with standard DB-061 pinnae.

Fig. 4.1 shows the resulting HRTF measured at the left ear (ear model DB-061) for a sound source at \(-40^\circ\) elevation and azimuth angles between 270° and 360°. These data should be comparable to the HRTF measured with the moulded pinna of subject LPO shown in Fig. 3.5 (notice that Fig. 4.1 shows the actual head-related transfer function, whereas in Fig. 3.5 the total pressure at the eardrum is represented). An overall comparison between the two figures shows that the most characteristic spectral features described in Chapter 3 are also present in the HRTF recorded using the standard DB-061 pinna even though the pinnae are different (see Appendix 1). It is particularly interesting that the spectral notches \( N_1 \), \( N_2 \) and \( N_3 \) are also present for the ear DB-061 and show an azimuthal dependency as described in Chapter 3. The main differences between the HRTFs for DB-061 and LPO is that the three notches have a
Chapter 4: Experimental HRTFs using a KEMAR with standard pinnae

slightly lower central frequency for the pinna DB-061, and N3 shows considerably more dependence on azimuth than for the LPO pinna.

The similarity between the results obtained with the standard DB-061 pinna and with the individualised moulded pinnae was expected because, after all, the standard pinnae are designed to generate average HRTFs (Burkhard and Sachs, 1975). However, so much similarity between both sets of results was welcomed in the sense that it validated our experimental method and the HRTF data obtained with it. It also meant that we could use the DB-061 HRTFs to investigate the elevation dependency of the spectral notches N1, N2 and N3 described in the previous chapter.

4.3.2 Elevation dependency

The HRTF data corresponding to sources placed at the same azimuthal angle but different elevation angles were plotted in the form of contour graphs in order to study the elevation dependency of the spectral notches. Results for the median vertical plane, lateral vertical plane, ipsilateral diagonal vertical plane and contralateral diagonal vertical plane are presented individually.

Median Vertical Plane (MVP). Fig. 4.2 shows the HRTF at the left ear of the KEMAR (ear model DB-061) for a sound source at 0° azimuth (source right ahead) and elevation angles between -40° and +90°. It can be seen that the centre frequency of N1 increases clearly with the angle of elevation, from around 6.3 kHz at -40° to 17 kHz at +90°. For elevation angles above +40°, the spectral shape is more like a gradual slope decreasing noticeably at frequencies above 10 kHz. The latter result is consistent with the spectra reported by Searle et al. (1975, figure 2). With regard to N2, it only appears at elevation angles below -10° and its centre frequency (10.5 kHz) does not show any particular trend with elevation. N3 is only seen below 0°. However, it looks as though its centre frequency follows a similar trend with elevation as N1, from 15.2 kHz at -40° to 16.3 kHz at 0°. Unfortunately, no data above 17 kHz was available to follow up this trend at higher elevation angles.

Lateral Vertical Plane (LVP). Fig. 4.3 shows the HRTF at the left ear of the KEMAR (ear model DB-061) for a sound source at 270° (source on the left) azimuth and elevation angles between -40° and +70° (with elevation steps of 10°). In
this case, the depth of NI has decreased with respect to the median vertical plane by approximately 10 dB for most elevation angles. However, notice the elevation dependency of NI is still observed despite the decrease in notch depth. (Notice how NI is deeper at elevations angles around 0°). Regarding N2, it is observed that its centre frequency increases with elevation, from 11.2 kHz at -40° to 17 kHz at +10°, following a trend similar to NI. This result is surprising since the elevation trend of N2 did not appear in the MVP. As for N3, it hardly appears in the LVP; Indeed, it slightly appears at 17 kHz at -40° elevation. The reason why N3 does not show in the lateral vertical plane within the frequency range considered is because the centre frequency of the spectral notches increases from the MVP to the LVP (as shown in Fig. 4.1). For instance, the centre frequency of N3 at -40° elevation is 15.3 kHz in the MVP, that is almost 2 kHz lower than in the LVP. A similar shift is observed for the centre frequency of N1 and N2 (see Fig. 4.1), although is important to notice that the higher the frequency the more acute the shift is.

Ipsilateral Diagonal Vertical Plane (see Chapter 1, section Angle conventions). Fig. 4.4 shows the HRTF at the left ear (model DB-06I) for a sound source at 315° azimuthal angle and elevation angles between -40° and +70°. Notice the source is at an intermediate azimuthal position between the median and the lateral vertical planes. Results (Fig. 4.4) reveal that N1, N2 and N3 are also clearly present in this vertical plane. N1 and N3 show a dependency on elevation similar to that in the median plane. Another interesting feature is that the centre frequency of N2 shows a clear trend to increase with the angle of elevation, just as it does in the lateral vertical plane. As pointed out above, the centre frequencies of N1, N2 and N3 are at an intermediate position between those in the lateral and median vertical plane.

Contralateral Diagonal Vertical Plane (see Chapter 1, section Angle conventions). Fig. 4.5 shows the HRTF at the left ear (model DB-06I) for a sound source at azimuthal angle of 45° (contralateral side) and elevation angles between -40° and +70°. It is very interesting to see that even though the source is at 45° on the contralateral side, N1 is still clearly present and its dependence on elevation is very similar to that in the other ipsilateral vertical planes described above. N2 is also clearly present, but its dependence on elevation is not so clear. It is also very clear that N2 is formed by two sub-notches N2a and N2b at 9.5 kHz and 11.3 kHz respectively. These two notches can be also easily observed in Fig. 4.1.
Figure 4.1. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimuthal angle, for a source at $40^\circ$ elevation. The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.2. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle, for a source in the median vertical plane (0\degree azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.3. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle, for a source in the lateral vertical plane (270° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.4. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle, for a source in the *ipsilateral diagonal vertical plane* (315° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.5. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle, for a source in the contralateral diagonal \textit{vertical plane} (45° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
4.3.3 Azimuthal dependency

The HRTF data corresponding to sources placed at the same elevation angle but different azimuthal angles were plotted in the form of contour graphs in order to study the azimuthal dependency of the spectral notches. Results are presented at $-40^\circ$, $0^\circ$ and $+40^\circ$ elevation individually.

$-40^\circ$ elevation. The azimuthal dependency of the HRTF at $-40^\circ$ has already been described above (see Fig. 4.1 and Chapter 3).

*Horizontal Plane (0° elevation).* Fig. 4.6 shows the HRTF results for the left ear (model DB-061) for a sound source at $0^\circ$ elevation and azimuthal angles between $270^\circ$ (source on the left) and $360^\circ$ (source right ahead). Only N1 appears consistently at this elevation, moving from 8.2 kHz at $0^\circ$ to 9.4 kHz at $270^\circ$. N1 is deeper and wider in the horizontal plane ($0^\circ$ elevation) than at $-40^\circ$ elevation. N2 is not visible at azimuth angles between $310^\circ$ and $0^\circ$; it only appears at 14 kHz at $310^\circ$ and moves to 16 kHz at $270^\circ$.

$+40^\circ$ elevation. Fig. 4.7 shows the HRTF for the left ear (model DB-061) for a sound source at $+40^\circ$ elevation and azimuthal angles between $270^\circ$ and $0^\circ$ ($10^\circ$ azimuth steps). In this plane the notches appear not to be present and the results show a slope-like shape gradually decreasing towards the high-frequency end. From Figs. 4.2, 4.3 and 4.4 it could be argued that the notch at 17 kHz may be N1.
Figure 4.6. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimithal angle, for a source in the horizontal plane (0° elevation). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Chapter 4: Experimental HRTFs using a KEMAR with standard pinnae

Figure 4.7. HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimuthal angle, for a source at ±40° elevation. The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
4.4 DISCUSSION and CONCLUSIONS

The HRIRs provided by Gardner and Martin (1994, 1995) have been analysed to obtain their corresponding HRTFs. The resulting HRTFs have been compared with the data presented in Chapter 3 for individualised moulded pinnae. The comparison has revealed many common spectral features. Among them are the characteristic spectral notches N1, N2 and N3. All three notches show an azimuthal dependency that is very similar in the case of the measurements made with the standard DB-061 pinna and those made with individualised moulded pinna described in Chapter 3 (particularly for subject LPO). Differences in the fine spectral characteristics must be related to differences in pinna shape. Differences in the actual centre frequency of the notches are likely to be related to size differences between the pinnae. In general, the larger the pinna, the lower the centre frequency of the spectral features. Notice that the DB-061 pinna (in particular the concha) is smaller than LPO’s pinna (see Appendix 1).

4.4.1 Azimuthal dependency

The results presented above show that the centre frequency of the most salient spectral notches are largely invariant as a function of the azimuthal angle of the source. This characteristic was also found in the HRTF measurements made with individualised moulded pinnae which were presented in Chapter 3. It must be pointed out, however, that spectral features at higher frequencies (e.g., N3) show a stronger dependency on source azimuth than those in the lower frequency range (e.g., N1). The reason for this behaviour is unclear and will be investigated in the following chapter. Another interesting finding is that the slight azimuthal dependency of the HRTFs occurs whatever the elevation angle of the sound source.

4.4.2 Elevation dependency

The HRTF data for the standard DB-061 pinna have been used to investigate the elevation dependency of the spectral notches N1, N2 and N3. Results have shown that the centre frequency of N1 increases rapidly with source elevation for angles
between $-40^\circ$ and $+40^\circ$. This result is particularly clear for sound sources in the median vertical plane (Fig. 4.2), but the data shown demonstrate that the same elevation-dependent behaviour occurs for sources in the lateral vertical plane and in the ipsi and contralateral diagonal planes. Furthermore, the azimuthal dependency of N1 at $-40^\circ$, $0^\circ$ and $+40^\circ$ elevation (shown in Figs. 4.1, 4.6 and 4.7 respectively) indicates that the elevation dependency of N1 is identical at any vertical plane within the frontal quadrant of the ipsilateral hemisphere. Even for sources in the contralateral hemisphere (Fig. 4.5) N1 shows a similar dependence with source elevation.

Evidence of an elevation-dependent notch in the median vertical plane has been given by Hebrank and Wright (1974) and Butler and Belendiuk (1977). Shaw and Teranishi (1968) and Bloom (1977) reported an elevation-dependent notch for sources in the lateral vertical plane. The results presented above show that the characteristic elevation dependency of N1 is not an exclusive feature of the median or the lateral vertical planes and, in fact, occurs at any vertical plane in the frontal hemisphere.

4.4.3 Generalisation of the results to other pinnae

In order to examine the generalisation of this result to other pinnae, the HRTF corresponding to the right ear (model DB-065, see Appendix 1) were analysed. Results for the median vertical plane are shown in Fig. 4.8; Results for the ipsilateral diagonal plane (i.e., $45^\circ$ azimuth) are given in Fig. 4.9 and those corresponding to the lateral vertical plane (i.e., $90^\circ$ azimuth) are shown in Fig. 4.10. There are obvious differences between the results obtained with the ear DB-061 and the ear DB-065. However, N1 is clearly present in both data sets. Moreover, the elevation dependency of N1 appears to be almost identical for both pinnae at the three vertical planes studied. This supports the generality of the results reported above for the ear DB-061 regarding the elevation dependency of N1 in any vertical plane. The generalisation of this interesting result to other pinnae is further supported by the fact that the response of the standard ears DB-061 and DB-065 (fitted onto a KEMAR with ear-canal simulators) is regarded as the average of the human ear (Burkhard and Sachs, 1975). However, one should always bear in mind that, in the same way that there are certain dissimilarities in the HRTFs measured with the two standard ears, spectral features specific to a particular subject may not be reproduced in the HRTFs measured using the standard ears DB-061 and DB-065.
Regarding N2 and N3, there is not an immediate correspondence between them (measured with DB-061) and the notches observed with the DB-065 ear. However, considering that the DB-065 pinna is larger than the DB-061 model (see Appendix 1 for a size comparison) one would expect the central frequency of corresponding pinna-related spectral features to be lower for the DB-065 pinna than for the DB-061 replica. This frequency shift would be accentuated for features at higher centre frequencies, i.e. for N2 and N3. This argument indicates that the notches at 9.5 kHz and 13 kHz observed at -40° elevation with the large DB-065 pinna (see Fig. 4.8) may correspond to N2 and N3 (centred at 10.4 kHz and 15.5 kHz respectively) measured with the small DB-061 pinna (Fig. 4.2).

Another possible interpretation of the data is that N2 in Fig. 4.2 (for the ear DB-061) is not a single notch, but, on the contrary, is comprised of two 'sub-notches' (N2a at 10.2 and N2b 11.5 kHz respectively). A closer look at the HRTF for the DB-061 ear for a source at -40° elevation, 0° azimuth supports this interpretation. The notches are so close together that appear as wider single notch (at 10.5 kHz) in our graphical representation. These two 'sub-notches' would correspond to the troughs at 9.5 and 13 kHz observed with the ear DB-065 at the same source location. Notice that the HRTFs for the DB-061 ear for a source in the contra-lateral diagonal plane (Fig. 4.5) shows that N2 splits into two separate notches similar to those observed with the DB-065 for a source in the MVP (Fig. 4.8). Whatever the interpretation, the elevation dependency of N2 is, unlike N1, different for the two ears, which indicates that this feature is more subject-specific.

With regard to N3, results are also inconclusive. Fig. 4.2 suggests that the dependence of N3 on the angle of elevation is similar to N1, at least for the source positions at which N3 is visible within the frequency range considered. This indicates that N1 and N3 may have the same physical cause, N3 being a higher order minimum of the fundamental N1 as suggested in Chapter 3. However, this result is not general since a similar conclusion can not be drawn from the results obtained for the ear DB-065. Spectral features above 12 kHz are, after all, very dependent on pinna size and shape.
4.4.4 Generalisation of the results to human HRTFs

The slight azimuthal variation of N1 is not peculiar to the KEMAR ears and has also been observed in human HRTFs. (Mehrgardt and Mellert, 1977, Figs. 11 and 12; Carlile and Pralong, 1994, Fig. 4; Pralong and Carlile, 1944, Fig. 3). In Carlile and Pralong's human HRTFs, there is a spectral notch (possibly corresponding to N1) whose central frequency varies from around 6 kHz when the source is in the MVP to 7.5 kHz for a source in the LVP at 0° elevation. Pralong and Carlile present human HRTFs which show a spectral notch (possibly N1) whose central frequency hardly deviates from 7 kHz when the source moves from the MVP to the LVP at 0° elevation. It is worth noticing that the HRTFs presented in those reports were recorded at 0° elevation and this may be the reason why N2 is either not present for all source azimuths or its azimuthal dependency is more variable than that observed in Fig. 4.6 for the KEMAR (ear DB-061).

The pronounced dependency of N1 with source elevation has been also observed in human HRTFs. It has been reported as a general feature in human HRTFs by Shaw and Teranishi (1968), Butler and Belendiuk (1977) and Bloom (1977). A similar elevation-dependent notch is described by Carlile and Pralong (1994) in their results for elevation varying sources in the MVP and the LVP.

4.4.5 Psychophysical implications

The results presented above regarding the dependency of N1 with source location may have important implications from a psychoacoustical point of view. The studies of Hebrank and Wright (1974) and Butler and Belendiuk (1977) concluded that N1 may be responsible for the sensation of elevation (or elevation judgement) of sound source in the MVP. Bloom (1977) demonstrated that a spectral notch whose central frequency increases with source elevation is perceived in the LVP. Up to now no evidence had been provided to demonstrate that similar elevation-dependent spectral notches appear for sources in any vertical plane. The results presented above provide that evidence. Therefore, even though no psychoacoustical experiments have been carried out to prove that indeed N1 is responsible for elevation judgements in any vertical plane, it is tempting to say that this may be the case. Some psychophysical
studies should be carried out to investigate this hypothesis formally. These studies could also investigate the relevance of N1 regarding azimuthal judgements.

4.4.6 The physical cause of the most salient spectral notches

The physical cause of N1, N2 and N3 has yet to be determined. Shaw and Teranishi (1968) concluded that N1 must be generated in the pinna as it appeared in their measurements under the blocked-meatus condition. More specifically, they speculated that it may be caused by interference between the two successive resonance modes generated in the concha, namely P2 and P3. Notice that their explanation was based on HRTF data for sources in the lateral vertical plane. We have shown above (and in Chapter 3) that P2 and P3 are mostly excited for sources in the lateral vertical. One would assume, therefore, that N1 would be deeper in the lateral vertical plane, where the interference effect should be maximal. The data presented above indicate that this is not the case and, in fact, N1 is deeper for sources in the median vertical plane.

Hebrank and Wright (1974) proposed a different explanation for N1. They speculated that N1 is caused by interference between the direct sound entering the ear-canal and the reflected sound from the posterior wall of the concha. Their explanation was suggested after their HRTF data for sources in the median vertical plane. Hebrank and Wright formulated their hypothesis on the basis that the extra distance travelled by the reflected wave within the concha generates interference at frequencies close to the centre frequency of N1. This theory might explain the presence of N1 in the median plane, for which reflections of the incident sound on the concha posterior wall are possible. However, it has been shown above that N1 is also present for sources in other vertical planes for which reflections on the concha posterior wall are not so obvious.

None of these two hypotheses has been formally proven. A attempt is made in the next two chapters to give a formal explanation to the observed spectral notches.
Figure 4.8. HRTFs measured at the right eardrum of a KEMAR (pinna model DB-065) as a function of the elevation angle, for a source in the median vertical plane (0° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.9. HRTFs measured at the right eardrum of a KEMAR (pinna model DB: 055) as a function of the elevation angle, for a source in the ipsilateral diagonal vertical plane (45° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Figure 4.10. HRTFs measured at the right ear drum of a KEMAR (pinna model DB-065) as a function of the elevation angle, for a source in the lateral vertical plane (90° azimuth). The HRTF data were obtained by analysing the HRIR database provided by Gardner and Martin (1995). See text for details.
Chapter 5: A physical model of the human concha

CHAPTER 5

A physical model of sound diffraction and reflections in the human concha

5.1 INTRODUCTION

This chapter is concerned with presenting an approximate physical model of the transfer function of the human concha. The model is used to investigate theoretically how the characteristic spectral notches described in the previous chapters are generated and how their dependence on source location (azimuth and elevation) comes about.

The main spectral features (peaks and notches) produced by the direction-dependent filtering of the pinna have been described in the previous chapters. The dependence of the external ear's transfer function on azimuth and elevation has also been well characterised. As explained in Chapter 3, the studies by Teranishi and Shaw (1968) and Shaw and Teranishi (1968) permitted identification of the physical cause of most of the spectral peaks measured at the eardrum position for incident flat-spectrum broad-band signals. However, the physical cause of some remarkable spectral notches remained unclear.

Clearly, if the incident undisturbed signal has a flat spectrum, spectral notches must be caused necessarily by destructive interference at the meatus entrance (Shaw and Teranishi, 1968). According to Batteau's theory of pinna reflections (Batteau, 1967), interference between the direct wave and the reflected wave from the whole of the pinna could be responsible for them. Hebrank and Wright (1974), however,

1 The work presented in this chapter has been accepted for publication in The Journal of the Acoustical Society of America under the title "A physical model of sound diffraction and reflections in the human concha," by Enrique A. Lopez-Poveda and R. Meddis. The paper is scheduled to appear in the issue of November 1996.
Chapter 5: A physical model of the human concha

hypothesised that reflections from the posterior wall of the concha alone may be responsible for the observed notch in the median plane. In an attempt to evaluate the validity of this statement, Hebrank and Wright regarded the human concha as a single-delay-and-add system where the elevation-dependent delay, \( \tau(\phi) \), between the direct sound entering the meatus and the reflected sound from the concha's posterior wall is related to the extra path, or path difference \( (d) \), travelled by the reflected wave. The relation is as follows: \( \tau(\phi) = d/c \). For sound sources in the median plane, each elevation angle would generate a particular value of \( d \) and thus a different delay. Since a single path difference is considered by this reflection theory, it implies that reflection occurs at a single point on the concha's posterior wall. The single-delay-and-add formulation provides a discrete collection of notch frequencies given by the expression \( f_n = n(1/2\tau(\phi)) \), \( (n = 1, 3, 5, \ldots) \).

Despite the simplicity of the single-delay-and-add model, its predictions indicate that reflections from the posterior wall of the concha may be responsible for the production of the elevation-dependent spectral notch in the median vertical plane. However, a satisfactory explanation to why a similar elevation-dependent notch has been also observed for incident sounds from other vertical planes (as shown in Chapter 4) can not be intuitively inferred from the mechanics of the concha reflections model explained above. For example, for incident sounds from the lateral vertical plane, the concha posterior wall is parallel to the direction of incidence and, therefore, reflections from the posterior wall of the concha into the meatus are not obvious. Despite this, it is observed that the centre frequency and the elevation-dependent behaviour of the notch in question are similar when the sound source is placed in the median and the lateral vertical planes as though they were caused by the same mechanism.

The aim of the work presented in this chapter is to investigate a theoretical explanation for some of the spectral features (peaks and notches) observed in the experimental head-related transfer functions presented in Chapters 3 and 4, in particular those caused by the concha. Special attention is dedicated to the characteristic elevation-dependent spectral notch, N1, described above. The basic hypothesis that guides our approach is that the elevation-dependent notch observed for incident sounds from both the lateral and median vertical planes are caused by cancellation between the direct wave entering the meatus and the sound reflected from the concha posterior wall. However, sound diffraction is necessary to explain how similar reflections on the
posterior wall of concha occur in both cases. In our view, the sound wave is diffracted (scattered within the concha) as it enters the concha cavity and, therefore, significant reflections must occur at an infinite number of points along the posterior wall for all source locations.

To test our hypothesis we have developed a model to simulate the spectral features related to the transverse geometry of the concha cavity. The model is inspired by the idea of concha reflections proposed by Hebrank and Wright (1974) but includes sound diffraction in the concha aperture as well as reflections on a large number of points on the concha posterior wall. Under certain assumptions and geometrical approximations, an attempt is made to develop the model in the frequency domain from simple physical principles. The predictions of the diffraction/reflection model are then tested against experimental measurements for an ideal diffracting/reflecting system with a spiral shape (Fig. 5.1b). The predictions of the diffraction/reflection model regarding the position of spectral notches are quantitatively compared with the predictions of the single-delay-and-add model for this simple geometry.

The proposed diffraction/reflection model is also applied to a more realistic three-dimensional concha-like shape (Fig. 5.3) in an attempt to simulate the spectral features (peaks and notches) generated by the tranverse dimensions of the concha. For the latter case, the diffraction/reflection model predicts a spectral notch which shows a dependence on azimuth and elevation similar to the notch N1 described in Chapters 3 and 4. Additionally, the model results suggest that secondary notches occurring at higher frequencies (namely N3) within the audible range are also caused by the same diffraction/reflection process in the concha.

5.2 THE DIFFRACTION/REFLECTION MODEL: THEORY

An exact expression for the transfer function of any acoustic system requires the wave equation to be solved (generally by numerical methods), with the right boundary conditions for the geometry of the system. Even for objects with simple geometry and well-determined acoustic characteristics, the process of solving the wave equation is non-trivial and computationally very expensive (Filippi and Dumery, 1969; Cassot, 1975; Martin and Rizzo, 1989). This is true particularly for an acoustic system such
as the human ear. For these reasons we have restricted our analysis to the human concha and opted for an approximate approach to obtain its transfer function. The approximations and assumptions made in our approach are listed below. Obviously, since some approximations are required, the final transfer function is inexact. However, the proposed approximated transfer function can help us understand a great deal about the mechanisms which generate some of the most characteristic spectral features observed in experimental head-related transfer functions.

5.2.1 General

Because the dimensions of the concha aperture are comparable to audible sound wavelengths, the incident wave will be diffracted as it enters the concha cavity. Hence, it will spread within the cavity following the rules of diffraction. The diffracted wave will, then, be reflected from the wall of the concha. It is convenient to express the total sound pressure at the ear canal entrance as the sum of the direct unreflected pressure from the concha aperture plus the reflected pressure from the concha reflecting wall, i.e.:

\[ P_T = P_U + P_R \] (5.1)

In the frequency domain, the transfer function of the concha can be expressed as the ratio of the total pressure at the meatus entrance over the value that the incident wave would have at the same point if the concha was not present. Mathematically, the transfer function of the concha can be written as follows:

\[ \frac{P_T}{P_0} = \frac{P_U + P_R}{P_0} \] (5.2)

As it will be explained below, the phase of the total pressure at the meatus entrance has a directional contribution which depends on the location of the sound. Therefore, the above transfer function should predict some of the spectral transformations introduced by the concha as a function of source position.
5.2.2 Theoretical approximations and assumptions

Both magnitudes, $p_U$ and $p_R$, clearly depend on the shapes of the concha reflecting wall and the concha aperture which are extremely complicated and vary substantially across subjects. In order to simplify the problem, a few general approximations and assumptions regarding the geometry of the system have been made. In addition, the general diffraction theories that are used in the proposed model do not apply specifically for apertures with dimensions much smaller than the wavelength. Therefore, some assumptions need to be made so that those theories can also be applied in our case. These general approximations and assumptions are:

(i) The ratio $p_U/p_0$ will be approximated to 1. This is not strictly true, but it may be justified by the fact that the distance from the concha aperture to the meatus entrance is very small compared to the minimum wavelength (~1.8 cm) of the considered sounds (1 Hz to 17 kHz).

(ii) It will be assumed that the concha entrance is a two-dimensional aperture on an infinitely large sound-opaque screen. This approximation is convenient because it allows the application of Kirchoff's theory of diffraction by an aperture (Ditchburn, 1976; Hecht, 1990) to specify the diffracted pressure at any point within the concha cavity (particularly, at points on the reflecting surface).

(iii) For frequencies below around 6 kHz the wavelength will be much larger than both the dimensions of the concha aperture and the distances between it and the concha reflecting wall. Therefore, for frequencies lower than 6 kHz, Kirchoff's diffraction theory is inappropriate to find the diffracted pressure at points within the concha cavity. Nevertheless, we have tentatively tested the diffraction/reflection model based on Kirchoff's theory for lower frequencies and found that, if approximation (i) is made, reasonable predictions are obtained even at these lower frequencies.

(iv) It is assumed that the incident pressure on each element of reflecting surface is given only by the diffracted sound impinging directly from the concha aperture. In other words, the contributions from other reflecting elements to the incident pressure on each reflecting differential have not been considered. The effects of this approximation on the predicted spectra will be discussed below.
If the sound source is placed at least one meter away from the concha entrance, the pressure wave reaching the concha entrance can be approximated by a plane wave of the form:

\[ p_o(\hat{r},t) = P_o \exp[j(\hat{k} \cdot \hat{r} - \omega t)] \]  \hspace{1cm} (5.3)

If we choose a system of co-ordinates such that its origin is defined at the centre of the meatus entrance (see Fig. 5.1a), the value of the undisturbed incident plane wave at this point reduces to:

\[ p_o = P_o \exp(-j\omega t) \]  \hspace{1cm} (5.4)

For convenience, we have chosen the centre of the meatus entrance as the observation point. A more rigorous approach would consider the effect that longitudinal and transverse modes of resonance within the ear-canal have on the experimental head-related transfer functions (Shaw, 1975).

Under approximation (i), the expression for the transfer function of the concha becomes:

\[ \frac{P_T}{P_o} = 1 + \frac{P_R}{P_o} \]  \hspace{1cm} (5.5)

In order to obtain a final expression for the concha transfer function, only the ratio \( P_R/P_o \) needs to be calculated. Unfortunately, \( P_R \) can not be derived in a trivial manner because the incident pressure on the reflecting wall is the diffracted pressure from the concha aperture. Therefore, the diffracted pressure at any point on the concha wall must be found before attempting to derive an expression for the reflected pressure at the meatus entrance.
Figure 5.1. (a) A two-dimensional view of the geometry of the problem and convention taken for the definition of the different variables involved. (b) Experimental set-up and dimensions of the reflecting spiral-shaped surface.
5.2.3 Diffracted pressure at any point on the concha reflecting wall

Under approximations (ii), (iii) and (iv), the incident pressure at any point \( \mathbf{r}_q \) on the concha reflecting wall (see Fig. 5.1a) can be calculated using Kirchoff's scalar diffraction theory (Braddick, 1965; Ditchburn, 1976; for a complete review on diffraction phenomena see Hecht, 1990) for a two-dimensional aperture on an opaque screen and an incident plane wave. According to Kirchoff's theory, each differential of surface in the concha aperture, \( ds_l \), can be seen as a secondary radiator emitting a spherical wave whose value at the point \( \mathbf{r}_q \) on the reflecting surface is given by the expression:

\[
dp\left(\mathbf{r}_q, t\right) = \frac{\mathbf{K}(\theta_{ql}, \psi_l, r_{ql}) P_0}{r_{ql}} \exp\left[j(kr_{ql} - \pi/2) + \varphi_l\right]\exp(-j\omega t) ds_l \tag{5.6}
\]

When the direction of propagation of the incident plane wave [Eq. (5.3)] is not normal to the concha aperture [i.e., when the angle \( \vec{k}, \vec{r}_l \) is different from 90 degrees], each secondary radiator on the aperture, \( ds_l \), radiates with an initial phase, \( \varphi_l \), given by:

\[
\varphi_l = j\vec{k} \cdot \vec{r}_l \tag{5.7}
\]

where \( \vec{r}_l \) is vector position of the \( l \)-th surface differential, \( ds_l \). Note that \( \vec{k} \) is the wave vector of the incident plane wave and, therefore, the direction of incidence (i.e., source position) is implicit in the value of its components.

For an incident plane wave, the *diffraction obliquity factor* \( \mathbf{K}(\theta_{ql}, \psi_l, r_{ql}) \), introduced by Kirchoff's theory is given by:

\[
\mathbf{K}(\theta_{ql}, \psi_l, r_{ql}) = \left[ \frac{1}{\lambda} \left( \frac{\cos \theta_{ql} - \cos \psi_l}{2} \right) + j \frac{\cos \theta_{ql}}{4\pi r_{ql}} \right] \tag{5.8}
\]

(see Appendix 2). It introduces the fact that each surface differential on the aperture, \( ds_l \), does not radiate with the same amplitude in all directions. The maximum
radiation occurs in the direction of propagation of the incident plane wave, i.e., when \( \theta_{ql} = \text{angle}(\overrightarrow{r_l}, \overrightarrow{n_l}) \) equals zero and \( \psi_l = \text{angle}(\overrightarrow{k}, \overrightarrow{n_l}) \) equals 180 degrees (note that \( \overrightarrow{n_l} \) is a vector normal to the aperture surface at the position \( r_l \) and, therefore, it is defined outward-pointing on the aperture).

The total diffracted pressure incident on any point \( \overrightarrow{r_q} \) of the reflecting wall is, therefore, obtained by integrating Eq. (5.6) over the whole of the concha aperture, \( S_A \), i.e.:

\[
P_D(\overrightarrow{r_q}, t) = \iint_{S_A} K(\theta_{ql}, \psi_l, r_q) P_o \frac{r_{ql}}{r_{ql}} \exp[j(k r_{ql} - \pi/2) + \varphi_l] \exp(-j\omega t) ds_l \quad (5.9)
\]

### 5.2.4 Reflected pressure at the meatus entrance

The next step is to find the pressure at the meatus entrance due to all the waves which have been reflected once. In this case, the same principle used to derive Kirchoff's expression for the diffracted pressure (Eq. 5.6) is used to obtain the reflected pressure, although with the appropriate boundary conditions (see Appendix 2). To illustrate the theory in a simple manner, each differential of reflecting surface, \( ds_q \), can be seen as a point reflector which emits a wave whose value at the meatus entrance is given by:

\[
dp_R = \frac{R(r_q, \gamma_q) P_D(\overrightarrow{r_q})}{r_q} \exp(jkr_q - j\omega t) ds_q \quad (5.10)
\]

where \( r_q \) is the distance from the reflector \( ds_q \) to the meatus entrance. The amplitude with which each \( ds_q \) radiates is proportional to the amplitude of the spatial part of the diffracted pressure at \( ds_q \); i.e., proportional to \( P_D(\overrightarrow{r_q}) \).

If total reflection is assumed, the proportionality factor, \( R(r_q, \gamma_q) \), can be expressed as (see Appendix 2):

\[
R(r_q, \gamma_q) = \frac{\cos \gamma_q}{4\pi} \left( -jk + \frac{1}{r_q} \right) \quad (5.11)
\]
where $\gamma_q = \text{angle}(\vec{r_q}, \vec{n_q})$ (see Fig. 5.1a). Notice that $R(r_q, \gamma_q)$ acts like a directional factor for reflection; it allows maximum reflection in the normal direction to the reflecting surface (i.e., when the angle $\gamma_q$ is zero) and it produces no reflected pressure from points at which $\gamma_q$ equals 90 degrees. If total reflection is not assumed, the form of $R(r_q, \gamma_q)$ will be slightly different (see Appendix 2) so that it includes the acoustic properties of the reflecting surface.

Substituting the value of $p_D$ into Eq. (5.10), operating and integrating over the whole reflecting surface, $S_R$, we can obtain the reflected sound pressure at the meatus entrance, which is given by the expression:

$$p_R = \int_S \int_S \frac{R(r_q, \gamma_q)K(\theta_{ql}, \psi_l, r_{ql})P_o}{r_q r_{ql}} \exp\left\{j[k(r_q + r_{ql}) - \pi/2] + \varphi_l\right\} \exp(-j\omega t)ds_q ds_l$$

(5.12)

5.2.5 The transfer function of the concha

Finally, substituting Eqs. (5.4) and (5.12) into Eq. (5.5) and operating, a final expression for the transfer function of the concha is obtained:

$$\frac{p_T}{p_o} = 1 + \int_S \int_S \frac{R(r_q, \gamma_q)K(\theta_{ql}, \psi_l, r_{ql})}{r_q r_{ql}} \exp\left\{j[k(r_q + r_{ql}) - \pi/2] + \varphi_l\right\} ds_q ds_l$$

(5.13)

It is important to notice that only the variables $\varphi_l$ and $\psi_l$ depend on the direction of incidence of the sound (i.e., on the source location). The rest of the variables only depend on the geometry of the diffracting/reflecting system.

An analytical solution of the above integral depends strongly on the shape of the concha which needs to be a regular mathematical function in order to express $ds_q$ and $ds_l$ in terms of $\vec{r_q}$, $\vec{r_l}$ and $\vec{r_{ql}}$. However, real concha shapes are irregular and variable among individuals. The numerical evaluation of the transfer function is, on
the other hand, more convenient and generalisable to almost any smooth concave shape given by any discrete collection of surface elements. A vectorial approach can be used to relate $r_q$, $r_l$ and $r_{ql}$, and also to find the values of $\theta_{ql}$, $\psi_l$ and $\gamma_q$ (see Fig. 5.1a).

5.3 VALIDATION OF THE THEORY

In order to test the formulation presented above for a simple geometry, the frequency transfer function of a spiral-shaped stainless-steel reflecting surface (see Fig. 5.1b) was measured. The predictions of the diffraction/reflection model for the same three dimensional spiral shape were then compared with the experimental data.

5.3.1 Experimental method

The impulse response of a spiral-shaped stainless-steel reflecting surface with a Bruel & Kjaer microphone (model 4134, 1/2" condenser mic.) placed at the position\(^2\) indicated in Fig. 5.1b was recorded for three elevations of the sound source (−45°, −17° and +35°). The measurements were made under quasi-free-field conditions and the recording time was set short enough to ensure that no reflected sound from any undesired reflecting surface was recorded. The expected degree of error in the angle measurements was ±3 degrees. The distance between the speaker and the microphone was fixed at 1.5 m for each measurement. Simultaneously, a second microphone with identical frequency response was used to measure the undisturbed sound. This second microphone was placed at the same relative position and distance from the speaker as the one inserted within the diffracting/reflecting spiral.

The impulse responses were measured using a super-pulse as the stimulus. A super-pulse was generated at each elevation using the super-pulse technique described in

\(^2\) The contour of the metallic reflector in the YZ plane is described by a collection of pairs $(y,z)$ which are expressed as $y = C \beta \cos(\beta + \pi/2)$ and $z = C \beta \sin(\beta + \pi/2)$, where $\beta$ is the angle (in radians) with respect to the positive Y axis and $C$ is a constant. In our case $C = 0.676$. Therefore, the metallic diffracting/reflecting surface is, in fact, a spiral of radius $r = C \beta$, rotated an angle $\pi/2$, where only the values $\pi \leq \beta \leq 2\pi$ are considered. In the experimental set-up the microphone was placed at the centre of the spiral.
Chapter 3. At all three elevations it was checked that the super-pulse "click" signals reaching both microphones, in the absence of the diffracting/reflecting spiral, had almost a flat spectrum (with variations between −1.5 and 1.5 dB) within the frequency range studied (1 kHz to 17 kHz). The spectrum of the impulse responses reaching both microphones was calculated automatically for each recording with the aid of a dedicated Fast Fourier Transform card implemented in a Masscomp MC5450 computer. When using the diffracting/reflecting system, the ratio of the spectra of the signals recorded at both microphones was taken to be the frequency transfer function of the diffracting/reflecting spiral.

5.3.2 Results

Fig. 5.2 shows the experimental results together with the predictions of the diffraction/reflection model when evaluated numerically for a collection of surface elements that describe the same three-dimensional shape and dimensions as those of the experimental diffracting/reflecting spiral (see Fig. 5.1b). (The computer code used for evaluating the diffraction/reflection model for this particular geometry is given in Appendix 3). For this particular geometry, only the diffracting effect of the frontal aperture (S₁ in Fig. 5.1b) was modelled [i.e., Sₐ = S₁ in Eq. (5.13)]. In other words, the effects of the open sides S₂ and S₃ of the metallic spiral were not considered. Since the reflecting surface was a rigid metallic one, total reflection was assumed (Landau and Lifshitz, 1959, p. 287; Frey et al., 1982, p. 168).

In the third panel of Fig. 5.2, the experimental results obtained at −45° are compared with the predictions of the model in two cases: (a) (dashed line) when the implemented model is the one presented in Eq. (5.13), and (b) (continuous heavy line) when the factor 1/rₖ is omitted in Eq. (5.13) (i.e., when attenuation of the diffracted waves with distance is not considered). From those results it is obvious that, regarding the central frequencies of the peaks and the notches, the model's predictions are the same in either case. On the other hand, regarding the amplitude of the spectral features, significantly better results are obtained when no attenuation of diffracted wave with distance is considered. Although this result is only presented for −45° elevation, the same occurs for the other locations tested. The actual explanation as to why this happens remains unknown to the authors. Possible reasons are given in the discussion section below. Perhaps, because the distances within the
diffracting/reflecting spiral are (< 6 cm.) smaller or comparable to the studied wavelengths (> 1.8 cm.), no attenuation of the diffracted wave with distance needs to be considered\(^3\). Consequently, in order to have a functional model it was decided to evaluate the model considering no attenuation of the diffracted waves with distance and the subsequent effects that this assumption produces in the general formulation\(^4\). The predicted results for \(+35^\circ\), \(-17^\circ\) and \(-45^\circ\) under this modification are shown in Fig. 5.2.

\(^3\) Stephens and Bate (1966) obtained correct results when describing the diffraction by a single slit even though they tacitly omitted the attenuation with distance of the diffracted waves. They justified their assumption by the correct results that it yielded.

\(^4\) A immediate consequence of not considering attenuation of the diffracted waves is that the imaginary part of Kirchoff's obliquity factor disappears from Eq. (5.8). This can be demonstrated noticing that, when attenuation with distance of the diffracted wave is not considered, the Green's function [Eq. (A2.2) in Appendix 2] for the diffracting surface becomes:

\[ g(\omega, R_o) = \exp(jkR_o)/4\pi \]

and, therefore, the gradient through the diffracting aperture is no longer as shown in Eq. (A2.3) but:

\[ \nabla_{\omega} g(\omega, R_o) = jk \cos \gamma_o \delta(\omega, R_o) \]

The resulting Kirchoff's obliquity factor after substituting this expression into Eq. (A2.1) and operating is equal to the real part of Eq. (5.8).
Figure 5.2  A comparison of the experimental data with the predictions of the single-delay-and-add approximation and the proposed diffraction/reflection model, at three elevations (−45°, −17° and +35°), for a metal spiral-shaped reflecting surface whose dimensions are shown in Fig. 5.1b. The thin continuous line shows the experimental results. The heavy line shows the predictions of the diffraction/reflection model when no attenuation with distance of the diffracted wave is considered. The heavy dashed line in the third panel (−45° elevation) shows the predictions of the diffraction/reflection model when attenuation with distance of the diffracted wave is considered. The arrows indicate the cancellation frequencies as predicted by the single-delay-and-add approximation for the metallic spiral.
5.3.3 A comparison between the performance of the single-delay-and-add approximation and the diffraction/reflection model

The arrows in Fig. 5.2 show the cancellation frequencies as predicted by the single-delay-and-add formulation for this reflector (Hebrank and Wright, 1974). According to the single-delay-and-add theory, for sounds incident from the median plane the cancellation frequencies are given by \( f_n = n \left( 1/2 \tau(\phi) \right) \) where \( n = 1, 3, 5, \ldots \), and \( \tau(\phi) = d/c \). In this case, the path difference between the direct and reflected waves, \( d \), was calculated as \( d = 2r \), where \( r \) is the distance from the microphone to the point on the reflecting wall at which the sound would reflect if it travelled within the cavity keeping its original direction (i.e., keeping its original elevation) — as done by Hebrank and Wright.

Table 5.1 shows the frequency position of the experimental minima and the predictions of both the single-delay-and-add approximation and the diffraction/reflection model. The square of the difference between the experimental and the modelled results is calculated for each elevation and each spectral notch as a quantitative estimate of the discrepancy between experimental values and the predictions of both models. For each elevation, the sum of the squared differences is considered to be an indication of the overall performance of the models; that is, the smaller the sum, the better the agreement between experimental and modelled results. The analysis was confined to spectral notches because, as opposed to spectral peaks, they are very narrow and, therefore, have a well defined centre frequency. The outcome of this analysis is that the diffraction/reflection model is more accurate at predicting the frequency position of each one of the examined spectral notches with the only exception of N3 at \(-17^\circ\) elevation. But even at this elevation the overall performance of the diffraction/reflection model is better than that of the single-delay-and-add approximation.
Table 5.1. Comparison between the central frequency of the experimental (Exp.) notches observed in the transfer function of the metallic spiral-shaped system (see Fig. 5.2) and the predictions of the single-delay-and-add (SDAA) approximation and the diffraction/reflection (DR) model. All values are expressed in kHz. The squared difference ($d^2$) between the experimental values and the predictions of both models is also shown. For every angle of elevation the sum of the squared differences is regarded as an indication of the overall performance of each model.

It is also remarkable that the single-delay-and-add approximation would predict higher order notches at frequencies $f_n = nf_1$, ($n = 1, 3, 5, ..., $), where $f_1$ is the frequency of the first notch. At $-45^\circ$, for example, the first experimental notch is placed at $f_1=3.3$ kHz and therefore, according to the previous formula, the higher order successive minima should be at: $f_3=9.9$ kHz, $f_5=16.5$ kHz and $f_7=23.1$ kHz. Experimental results show that these minima actually occur at $f_3=7.6$ kHz, $f_5=12.3$ kHz and
\( f_r = 16.4 \) kHz. These figures clearly indicate that the single-delay-and-add approximation is inadequate to predict not only the absolute position of the spectral minima but also the relative position between them.

**Two-dimensional approximation**

In order to find an approximation for two-dimensional waves, the diffraction/reflection model was also evaluated for the two-dimensional projection of the metal spiral (i.e., for a contour such as the one shown in Fig. 5.1a). The surface integrals over \( S_A \) and \( S_R \) become, therefore, line integrals under this approximation. In order to get almost identical results for both the two-dimensional approximation and the real three-dimensional evaluations it was found that the reflection obliquity factor in the two-dimensional case needed to be \( R(r_q, \gamma_q) = \cos \gamma_q \) instead of the theoretical expression given in Eq. (5.11). This finding will be used when applying the model for the realistic concha contour (see below).

### 5.4 Predictions of the Diffraction/Reflection Model for a Realistic Concha Shape

In this section, the predictions of the diffraction/reflection model for a realistic concha shape are compared against the experimental head-related transfer functions described in Chapters 3 and 4.

The diffraction/reflection model was evaluated for an approximated concha shape with a geometry such that the spectral features related to the tranverse dimensions of the human concha should be predicted. (The computer code used for evaluating the diffraction/reflection model for the realistic concha shape is given in Appendix 3). In order to generate such a shape, the inner contour (Fig. 5.3a) of a moulded concha was measured from the centre of the ear-canal entrance to its posterior wall. An approximate three-dimensional shape was then generated by projecting the contour to a vertical plane oriented at a protrusion angle (Burkhard and Sachs, 1975) of 158° with respect to the frontal part of the median sagittal plane (Fig. 5.3b). The resulting geometry is an approximate concha cavity defined by the lateral reflecting wall, \( S_L \), the posterior reflecting wall, \( S_P \), and the diffracting aperture, \( S_A \) (see Fig. 5.3). The
dimensions of the concha breadth (≈1.8 cm.), concha depth (≈1.33 cm.) and protrusion angle (158°) in the approximate concha are close to the average values for a male ear reported by Burkhard and Sachs (1975).

Although reflections from the lateral wall are also considered in the general formulation of the diffraction/reflection model, their contribution to the reflected pressure at the meatus entrance is zero for the approximate concha considered in this case. The reason is that $\gamma_q = 90^\circ$ for all surface elements in $S_L$ and, therefore, the factor $\cos \gamma_q$ in the reflection obliquity factor [Eq.(5.11)] becomes zero over the lateral wall. Consequently, the integration over the reflecting surface in the diffraction/reflection model was evaluated over the posterior wall only. The evaluation of the diffraction/reflection model under these conditions should predict the spectral features related to the transverse dimensions of the concha.

The bottom panel of Fig. 5.4 shows the predictions of the diffraction/reflection model when it is applied to the approximate concha shape assuming a sound source in the median vertical plane at elevation angles between $-40^\circ$ and $+90^\circ$. The top panel of Fig. 5.4 corresponds to the experimental data shown in Fig. 4.2. The experimental data is reproduced here to facilitate the comparison with the model predictions. Similarly, the bottom panel of Fig. 5.5 shows the predictions of the diffraction/reflection model when it is applied to the approximate concha shape assuming a azimuth-varying source at $-40^\circ$ elevation. The top panel of Fig. 5.5 shows the experimental data corresponding to the same source locations. The experimental data is reproduced from Fig. 4.1 which is shown here again to facilitate the comparison with the model predictions.
Figure 5.3. (a) The inner contour (lateral view or YZ projection) of a real concha which was used to generate the approximate 3D-concha shape for which the
diffraction/reflection theory has been evaluated. (b) Top view (or XY projection) of the
approximate three-dimensional concha. Note that the protrusion angle is 158° and that
the concha cavity is effectively described by two reflecting surfaces ($S_L$ and $S_P$) and by
the diffracting aperture $S_A$. The axes units are given in cm. and express to scale the real
dimensions of the concha.
Figure 5.4. Top: HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the elevation angle of a sound source placed in the median vertical plane. This figure is a reproduction of Fig. 4.2. Notice that the data include the effect of all the body parts which modify the incident sound before it reaches the eardrum; i.e., torso, head, and the whole of the external ear. Bottom: Predictions of the diffraction/reflection model for a realistic concha shape (see main text for details). Only the spectral features associated with the tranverse dimensions of the concha are plotted in this panel.
Figure 5.5 Top: HRTFs measured at the left eardrum of a KEMAR (pinna model DB-061) as a function of the azimuth angle for a sound source placed at $-40^\circ$ elevation. This figure is a reproduction of Fig. 4.2. Bottom: The predictions of the diffraction/reflection model for a realistic concha shape (see main text for details).
As it was found for the metal reflecting spiral, the amplitude of the predicted spectral features (peaks and notches) is very much smaller than the experimental values. In order to improve the predicted amplitudes, attenuation with distance of neither the diffracted nor the reflected waves was considered in this case. What is more, although the model was applied to a three-dimensional reflecting surface, the two-dimensional reflection obliquity factor, \( R(\gamma_q, \gamma_q) = \cos \gamma_q \), was used so that the amplitude of the reflected wave was larger. For the realistic concha, therefore, the diffraction/reflection model was evaluated using these modifications, whose theoretical implications will be discussed below. Moreover, the model was evaluated assuming total reflection.

The diffraction/reflection model predicts the location of notches N1 and N3 and the dependence of their central frequency on the elevation and azimuthal angles. With the modifications explained above, the depth of N1 and N3 is comparable to the experimental values for most azimuths and elevation angles between \(-45^\circ\) to \(+45^\circ\). In the median plane, however, the predicted depth of N1 (\(-9\) dB maximum) is still smaller than the experimental values (which are of the order of \(-15\) to \(-20\) dB maximum). The modelled data also anticipate that N1 is deeper at elevations around 0° for any vertical plane. This prediction coincides with the experimental data shown in this paper and the data reported by Shaw and Teranishi (1968) and Bloom, (1977). For elevations higher than \(+45^\circ\), an individual analysis of the predicted spectra shows that the modelled N1 appears at higher frequencies (around 16 kHz for \(+90^\circ\)) and is much wider than at low elevations. This may be why the experimental spectra appears like a gradually decreasing slope for elevation angles close to \(+90^\circ\), as pointed out in Chapter 4. Even with the modifications explained above, the depth of the predicted notches at elevations higher than \(+45^\circ\) is very much smaller than the experimental values (see Fig. 5.5) for any azimuthal angle.

The model does not predict the experimental notch N2. Instead, it predicts a resonance between 10 and 13 kHz, which corresponds to the transverse mode of resonance of the concha —or vertical mode— (Shaw, 1975).
5.5 DISCUSSION

The experimental HRTFs are more irregular than the model predictions. A reason for this dissimilarity is that, while the modelled results only reflect the spectral features generated by the posterior wall of a simplified concha, the experimental data also includes the effects of the concha depth, head, torso, pinna flange, helix, tragus, ear canal and other anatomical parts of the human pinna. All of these parts also contribute to the final spectra recorded at the eardrum (Teranishi and Shaw, 1968; Shaw, 1975). Additionally, the theory has been evaluated assuming ideal total reflection even for the real concha. Obviously, this is not the real case. Unfortunately, no precise data about the acoustic properties of the material used to make the manikin’s ears were available when carrying out these studies.

For sounds incident from the median plane, the diffraction/reflection model supports Hebrank and Wright’s speculation that the notch N1 is caused by cancellation between the direct sound entering the meatus and the reflected wave from the posterior wall of the concha. However, in the diffraction/reflection theory reflections from the whole of the concha’s posterior wall have been considered. Therefore, the dependence of N1 and N3 on elevation must be the result of the asymmetry of the reflecting wall with respect to the meatus entrance together with the weighting of the amplitudes for the multiple delays which is given (as explained in the theory) by the obliquity factors for diffraction and reflection.

The most important feature of the diffraction/reflection theory is, perhaps, that it predicts and explains why similar spectral features (in particular N1 and N3) have been observed when the sound source is placed in any vertical plane within the frontal hemisphere. In other words, it predicts the virtual independence of the central frequency of N1 and N3 on the azimuthal position of the sound source. The explanation is as follows. Whatever the incident direction (source position), the diffraction process scatters the sound within the concha cavity. Consequently, reflections on the posterior wall of the concha occur for any incident direction of the sound (although the diffraction obliquity factor encourages sounds coming from the frontal part of the ipsilateral hemisphere). Interference between the direct and reflected waves at the meatus entrance generates, therefore, a similar spectral pattern whatever the azimuthal angle of the sound source. Even for sources placed in the frontal part of
the contralateral side the same mechanism applies if we accept that the sound "creeps" around the head entering the contralateral concha at *approximately* the same elevation angle as if the source was in the ipsilateral hemisphere. This mechanism would explain, therefore, why N1 is also present for a sound source in the contralateral diagonal vertical plane as shown in Fig. 4.5. However, when the sound source is off the median plane, the effect of the head must also be taken into account to evaluate the amplitude of the individual frequency components of the signal reaching the concha aperture.

The reason N1 should be deeper at elevations around 0° is also explained by the diffraction/reflection model. From Eq. (5.13) it is obvious that deeper notches must occur for elevation angles, φ, at which the normalised amplitude of the reflected wave is closer to unity. The only direction-dependent factor in the amplitude of the reflected wave is the diffraction obliquity factor, $K(\theta_{qi}, \psi_i, r_{qi})$ — Eq. (5.8). The other factors only depend on the geometry of the problem. Therefore, maximum reflected amplitudes are obtained for elevation angles at which $K$ is maximal. From Fig. 5.1a it can be seen that $\cos \theta_{qi}$ is direction-independent and $\cos \psi_i = -\cos \phi$. Consequently, at 0° elevation $\cos \psi_i = -1$ and, therefore, $K(\theta_{qi}, \psi_i, r_{qi})$ reaches its maximum value with respect to φ.

Regarding the amplitude of the simulated spectra, certain modifications (listed above) have been introduced in the model in order to generate spectral features whose amplitudes are comparable to the experimental values. The main effect of introducing such modifications is an overall increase of the amplitude of the reflected wave (i.e., the amplitude of $p_R/p_o$) which, therefore, generates higher peaks and deeper notches at those frequencies at which the phase of the reflected wave is close to 0 and π radians respectively. It is important to notice that those modifications do *not* modify the dependence of the central frequency of the predicted features on the incident direction of the sound which is included in the model by the initial phase $\psi_i$ and the angle $\theta_{qi}$.

A robust theoretical argument which explains why those modifications are required is not known by the authors but it is most likely a consequence of the approximations made for developing the diffraction/reflection transfer function. It is important to notice that the single-reflection approximation (iv) yields a reflected wave whose

---

5 It should be noticed that, although the real system is three-dimensional, the two-dimensional description given in Fig. 5.1a is enough to illustrate the dependence of K on elevation.
amplitude is smaller than the value it would have if the approximation were not assumed (i.e., if multiple reflections were considered). It may also be that, in the case of the real concha, the size of the diffracting and reflecting surfaces and the distance between them is so small compared to the wavelength that attenuation with distance does not need to be considered at all.

Even with the adjustments explained above, results have shown that for elevation angles greater than around +40°, N1 is less deep in the model than in the experimental data. This may be partially explained because Kirchoff's theory of diffraction does not work very accurately for angles of incidence far away from the normal to the diffracting aperture (Braddick, 1965; p. 176). In addition, in the real case, the pinna flange must cause some important "shadow effect" for sounds coming from high elevations.

The diffraction/reflection model presented here cannot predict the minimum at 10.5 kHz (N2). It may well be caused by an anatomical part of the real pinna not considered in the model or it may be related to other physical phenomena than those related to the tranverse dimensions of the concha. The fact that it is right in the middle of the predicted vertical resonance of the concha indicates that it may be caused by the presence of the crus helias and the fossa in the real concha, as Shaw (1975) observed. This possibility is investigated experimentally in the next chapter.

5.6 CONCLUSIONS

An approximate physical model of sound diffraction and reflections in the human concha has been proposed. The model constitutes, to our knowledge, the first attempt ever to include diffraction in the formulation. Results are promising regarding the predictions of the centre frequency of some of the most characteristic spectral notches. However, some modifications in the theoretical formulation of the diffraction/reflection model have been required in order to optimise the amplitude of the predicted spectral features. These modifications only increase the depth of the predicted minima and do not change their centre frequencies or their dependence on source location. The need

---

6 Only at distances large compared with the size of the source and with the wavelength is the flow of energy proportional to the inverse square of the distance from the source (Ditchburn, 1976; p. 25).
for these modifications is likely to be a direct consequence of the approximations made when developing the diffraction/reflection model of the concha.

Although the proposed diffraction/reflection model accounts for some of the spectral features related to the tranverse dimensions of the concha, it is only an approximation. Further investigation needs to be done in two different directions in order to predict a complete functional model of the human concha. Firstly, a general formulation which needs no major assumptions and approximations should be developed. Secondly, the improved formulation should be adequate to explore in a unified manner the spectral features originated by both the depth of the concha and its tranverse dimensions, and by the interactions between them. This could only be achieved by solving, for each source location, the wave equation with the right boundary conditions for an acoustic system such as the concha and, perhaps, the ear. Solving the wave equation for such a complicated system would require the acoustic characteristics (e.g., acoustic impedance) of the external ear to be known as well as an accurate geometrical description of the external ear and/or the pinna replicas. The latter problem can be solved by digitising a pinna replica.

The diffraction/reflection model fails to predict the spectral notch N2. It has been argued that it may be related to an anatomical feature not included in the approximate concha shape for which the model has been evaluated. The next chapter is concerned with examining the overall contribution of the most significant anatomical features of the external ear to the experimental head-related transfer functions. Particular attention is given to investigate whether N2 is related to any particular anatomical feature of the external ear.
CHAPTER 6

The contribution of individual anatomical features of the pinna to experimental head-related transfer functions

6.1 INTRODUCTION

In Chapters 3 and 4 new experimental data in the form of head-related transfer functions were presented. The data were used to examine the elevation and azimuth dependency of salient spectral features. The physical cause of spectral peaks was explained after the studies by Shaw and Teranishi (1968), Teranishi and Shaw (1968) and Shaw (1975). Additionally, it was pointed out that the cause of important characteristic spectral notches was unclear. An approximate physical model of the human concha was presented in Chapter 5. The predictions of this model suggest that the origin of the characteristic elevation-dependent spectral notches, N1 and N3, is to be looked for in the concha. However, the model does not account for a major spectral trough at 10 kHz (N2). Since the model was evaluated for a relatively simple (although realistic) concha shape, it was argued that N2 could be caused by other anatomical features not considered in the model.

This chapter is concerned with investigating the individual contribution of some of the anatomical features of the external-ear (Fig. 2.4) to experimental head-related transfer functions (HRTFs). There are two possible interpretations to the spectral features observed in head-related transfer functions. On the one hand, it may be that a one-to-one relationship exists for all source locations between specific pinna-based spectral features and particular anatomical structures of the external ear. This interpretation is suggested by the evidence that the spectral peaks described in the previous chapters can be directly attributed to specific cavities of the external ear. On the other hand, it
seems reasonable to think that individual anatomical elements of the pinna must interact primarily with those frequency components whose corresponding wavelengths are comparable to the physical dimensions of the anatomical element (see Chapter 2, Background). This suggests that particular anatomical elements of the pinna must determine the spectral shape of the HRTFs around the same frequency region for all source locations. The specific frequency region would be related somehow to the dimensions of the anatomical element in question.

One of the aims of the work presented in this chapter is to investigate the accuracy of these two possible interpretations. In other words, to investigate whether a one-to-one relationship can be established between specific pinna-based spectral features and particular anatomical elements of the pinna, or, on the contrary, specific anatomical elements influence the HRTF shape around the same frequency range for all source locations. If the first statement was true, we should be able to find out the anatomical feature responsible for N2 and check the validity of the diffraction/reflection model from a quantitative point of view.

The experiments presented here were designed to investigate particularly the effects of the flange cavities (fossa and scapha), the flange itself, the concha cavity, the crus helias and the tragus (Fig. 2.4). A KEMAR and an individualised moulded pinna (left LPO ear, see Appendix 1) are used. We start by measuring the head-related transfer function using the unaltered ear. The pinna is then altered either by filling in the flange cavities with pinna-moulding material (Otoform) or by gradually cutting off the anatomical parts whose effect we are interested in studying. Head-related transfer function measurements are produced at each stage for three different elevation angles \((-45^\circ, 0^\circ \text{ and } +45^\circ)\). At each elevation, measurements were made at 10 azimuthal positions (in steps of \(10^\circ\)) between the median and the lateral vertical planes.

An improved version of the super-pulse method presented in Chapter 3 was employed to make the HRTF recordings. As explained in Chapter 4, the Masscomp computer initially used to make the HRTF measurements was no longer functional. Therefore, the new super-pulse code was re-written and adapted to be run in an Indy Silicon Graphics work-station. The main features of the new super-pulse code and its performance are also explained below.
Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs

The results obtained are discussed in the context of previous similar studies. The compatibility of the results obtained with current psychophysical knowledge is also discussed.

6.2 EXPERIMENTAL METHOD

6.2.1 Experimental set-up and equipment: An improved super-pulse method

The super-pulse method explained in Chapter 3 was employed for measuring the experimental HRTFs. However, as explained above, a fatal failure of the Masscomp computer used in previous experiments (reported in Chapter 3) prevented us from making the HRTFs recordings using the Masscomp and its computer code. Consequently, the super-pulse code had to be re-written and adapted to be run using the audio ports of an Indy Silicon Graphics work-station. The new super-pulse code was written in LUTEar format and, therefore, makes use of some of the analysis and utilities functions available in the LUTEar library. For instance, the frequency analysis of the recorded impulse responses is done with the fast Fourier transform (FFT and inverse-FFT) functions available in LUTEar. These routines were adapted to LUTEar from Press et al. (1992).

We took advantage of having to re-write the super-pulse code to improve the efficiency of the method. The result is that the new super-pulse method is more efficient in terms of the time required for making a single-location HRTF measurement. This is achieved as follows. As explained in Chapter 3, an average response of the delivery/recording system needs to be measured prior to generating the super-pulse. In the old super-pulse code, this average response was calculated from the responses to a specified number (N) of individual Dirac pulses. This means that the process of delivering and recording the clicks was done N times, one per click. In the new super-pulse code, a single train with N Dirac pulses is delivered and recorded just once which speeds up the process enormously. The separation between two successive clicks in the train was fixed to 100 ms. This is an important parameter that needs to be large enough so that echoes associated with one click do not interfere with the impulse response associated with the next click in the train. The recorded signal is then
appropriately divided into N intervals (one for every click of the train) with a duration of 100 ms. The average response of the system is then calculated as the average across those N intervals. The average signal can then be cut shorter to ensure that no echoes are included as part of the impulse response of the system. In our case, the average signal was shortened from 100 ms to 50 ms. This guarantees that echoes from reflecting surfaces further than approximately 8.5 m are not considered (c = 340 m/s). This is the order of distances to be expected in quasi-free-field conditions. Of course, reflections must be kept to a minimum as they can not be compensated for even with the super-pulse method.

The same idea was applied for making the actual HRTF measurements with the super-pulse previously generated. In this case, however, the train of Dirac pulses was replaced by a train of super-pulses.

Fig. 6.1 shows the average system's response to a train of forty Dirac pulses compared with the average system's response to a train of forty identical super-pulses. The super-pulse was generated as explained above. In this particular case, the super-pulse was generated indoors in a non-empty room\(^1\). The room dimensions were 6x7.5 m and it was 3 m height. The delivery/recording system was placed approximately in the centre of the room. The speaker and the microphone were both at 1.65 m from the floor (i.e., the source was at 0° elevation). In these conditions, there must have been important reflections from the objects in the room and the room walls. These reflections were not eliminated as the duration of the super-pulse was fixed at 50 ms (see above). Additionally, the ambient noise in the room was typical of a computer laboratory with 9 computers switched on. No serious attempt was made to reduce the ambient noise, as it should be cancelled by the averaging process included in the super-pulse method. There is no doubt that these experimental conditions are far from ideal. However, Fig. 6.1 shows that even in this case the new method is able to generate a super-pulse which produces an average system's response with variations of less than ±2 dB between 1 and 2 kHz and less than ±1.5 dB above 2 kHz (our study was limited to frequencies between 1 and 17 kHz). As explained in Chapter 3, these variations are acceptable for our purposes of measuring head-related transfer functions.

\(^1\) It was actually generated in the main room of the Speech and Hearing Lab., Dept. of Human Sciences, Loughborough University.
The rest of the experimental equipment used in these experiments was the same as that described in Chapter 3 (Fig. 3.1). It is worth mentioning, however, that the Radio Spares (RS) speaker was used to deliver the sound at all three elevations (−45°, 0° and +45°) at which measurements were made. It was explained in Chapter 3 that the KEF speaker had been found to generate a better super-pulse at low elevation angles (−40°) whereas the RS speaker performed better at 0°. We took advantage of these experiments to carefully investigate the quality of the super-pulse generated with both speakers at different elevations. It was discovered that the super-pulse obtained with the RS speaker is equally good at all elevations (and better than the one obtained with the KEF speaker) as long as the speaker faces the microphone (or the KEMAR) directly and not at an oblique angle.
Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs

Figure 6.1. Spectra of the average response of the system to 40 Dirac pulses (thin line) and 40 super-pulses (thick line)—see text for details—. Note that the response of the system to the super-pulse is almost flat with maximum variations of ±1.5 dB between 2 and 17 kHz.
6.2.2 Procedure

Head-related transfer functions were measured with a B&K microphone (see Chapter 3) at the left eardrum of a KEMAR for sound sources at three elevation angles, $-45^\circ$, $0^\circ$ and $+45^\circ$. For every elevation, measurements were made every $10^\circ$ for azimuthal angles between the median and the lateral vertical planes ($0^\circ$ and $270^\circ$ azimuth respectively). The distance between the speaker and the ear-canal entrance was fixed to 1.4 m. The KEMAR wore neither clothes nor a wig and was fitted with Knowles ear-canal simulators (model DB-100) and the two neck rings during all experiments.

A new super-pulse stimulus was generated with the improved super-pulse method for each one of the three elevation angles. Because of the lack of an anechoic chamber, the HRTF measurements were made indoors in an ample room (see above). Our intention was to study the frequency range between 1 and 17 kHz, in order to be consistent with the experiments carried out in Chapter 3. The experimental conditions made it virtually impossible to obtain a super-pulse that generates a system's response with variations of less than $\pm 1.5$ dB for frequencies below 2 kHz. This was obviously caused by the number of reflections present in the room for which even the super-pulse method can not compensate. However, it was relatively easy to obtain super-pulse signals which produce a quasi-flat system's response (with variations of less than $\pm 1.5$ dB) for frequencies above 2 kHz. In any case, the non-flatness of the response below 2 kHz does not interfere with the purpose of these experiments because the pinna-based spectral features that are to be examined only appear at frequencies well above 2 kHz (see Chapter 3 and Shaw and Teranishi, 1968).

Five sets of HRTF measurements were produced at each one of the 30 positions tested. The first set was obtained for the KEMAR fitted with the left individualised moulded pinna LPO (shown in Fig. 6.2a and also in Appendix 1). The right infill (also shown in Appendix 1) replaced the right ear at all times. For the second set of measurements the flange cavities, particularly the fossa and the scapha, were filled with the same Ottoform compound (see Chapter 3, section 3.2) used to produce the pinna replica. Special care was taken so that the final filling did not increase the natural size of the concha or the flange itself. The resulting ear piece is shown in Fig. 6.2b. This condition will be referred to as the 'flange-filled' condition. The third set
of measurements was obtained with the LPO pinna from which the flange had been cut off. The resulting ear piece can be seen in Fig. 6.2c and, basically, consists of the concha alone. This condition will be referred to as the 'no-flange' condition. For the fourth set of measurements, the crus-helias (see Fig. 2.4) was removed from the concha used in the previous set. The resulting ear piece is shown in Fig. 6.2d and will be referred to as the 'concha-no-crus-helias' condition. Finally, for the fifth set of measurements the tragus was removed from the ear piece used in the previous set. The resulting ear piece is shown in Fig. 6.2e. This condition will be referred to as 'concha-no-tragus' condition.

As suggested in Chapter 3, a 'free-field' reference level was measured at each elevation angle so that the resultant HRTF data could be presented in the form of transfer functions (re free-field) instead of using absolute levels. By 'free-field' level it must be understood the level measured at the B&K microphone placed in the experimental room at the position of the centre of the KEMAR's head.

---

2 Notice that, since the experiments were carried out in a large standard room, the term 'free-field' is inexact.
Figure 6.2 (Continues in the next page). (a) Unaltered LPO left pinna (see also Appendix I). (b) 'Flange-filled' ear-piece. (c) 'No-flange' ear-piece. Notice that the flange has been removed. (d) 'Concha-no-crus-helias' ear-piece. Notice that the crus helias (see Fig. 2.4) has been removed from the 'no-flange' ear-piece shown in (c). (e) 'Concha-no-tragus' ear-piece. Notice the tragus has been removed from the ear-piece shown in (d).
Figure 6.2. continued (see legend in previous page).
Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs

6.3 RESULTS

Figs. 6.3, 6.4 and 6.5 show the resultant HRTFs at -45°, 0° and +45° elevation respectively. For each elevation, results for the five different conditions are shown in five adjacent panels from left (unaltered pinna) to right ('concha-no-tragus') in order to facilitate the comparison between different conditions. The data are represented as transfer functions (re 'free-field'). The x-axis represents the azimuthal angle, which extents from 270° to 360° for each condition. (Notice that 360° azimuth is equivalent to 0° azimuth). 'T' in the x-axis stands for 'transition' between two different conditions. For graphical representation purposes, T was assigned zero values. The information at T does not represent any real data and, therefore, must not be interpreted as part of the experimental HRTFs.

6.3.1 Unaltered pinna

-45° elevation. The results at -45° elevation for the unaltered-pinna condition (Fig. 6.3) should be comparable to those presented in Chapter 3 (Fig. 3.5a) for azimuthal angles between 270° and 360°, since they correspond to the same pinna (LPO) and similar source locations. This comparison is made below in order to examine the consistency of the results obtained with the two methods and under different experimental conditions. However, before making any comparison it must be pointed out that the elevation angle was different in both experiments (-45° for the results presented in Fig. 6.3, and approximately -40° for the results presented in Fig. 3.5a). Additionally, it should be noticed that Fig. 3.5a shows the absolute level at the eardrum position, whereas a transfer function (re 'free-field') is represented in Fig. 6.3. The latter remark is particularly significant since it may lead to misinterpretation of the data under the graphical representation used to display the data (see below).

Bearing the above considerations in mind, a comparison between the two data sets (Fig. 6.3 and Fig. 3.5a) shows the following:

- The ear-canal resonance, at around 2.8 kHz (P1), and the depth resonance of the concha, at around 4.6 kHz (P2), are unmistakably present in both Figs.
The first spectral notch, N1, described in Fig. 3.5a at around 7.2 kHz, is also present in Fig. 6.3 at roughly the same frequency. However, N1 appears to be shallower in Fig. 6.3 than in Fig. 3.5a for azimuth angles close to the lateral vertical plane (270° azimuth) and deeper for azimuth angles close to the median plane (360° = 0° azimuth).

The third and fourth peaks, P3 and P4, are also clearly present in both Figs. (P3 is centred at approximately 10.5 kHz at 320° azimuth, and P4 at around 13 kHz at the same azimuth).

The second notch, N2, is also clearly visible in both Figs. However, N2 appears at a slightly higher frequency in Fig. 6.3 than in Fig. 3.5a. It moves from 11.8 kHz in the median plane to 12.75 in the lateral plane in Fig. 6.3, whereas it goes from 10.5 kHz in the median plane to 12.5 kHz in the lateral plane in Fig. 3.5a.

Regarding N3, it appears more clearly in Fig. 6.3 than in Fig. 3.5a. Notice the strong dependency of N3 on the azimuthal angle in Fig. 6.3 (going from 14 kHz at 0° azimuth to 16.5 at 270°). N3 is not so distinctively present in Fig. 3.5a. However, it can be seen at roughly the same frequency.

The main difference between the two Figs. is that a clear notch (labelled N1b) appears between N1 and N2 in Fig. 6.3, which was not present in Fig. 3.5a at least for sources away from the median plane. The centre frequency of N1b changes from approximately 8.3 kHz for a source in the median vertical plane to 10 kHz for a source in the lateral plane. This notch was not specifically noticed in the description of Fig. 3.5a in Chapter 3. It was only after N1b was clearly seen in Fig. 6.3 that we noticed that it also appears in Fig. 3.5a at roughly the same frequency but only for azimuths close to the median plane. However, N1b does not appear at all azimuths in Fig. 3.5a, as it does in Fig. 6.3.

In summary, the results obtained in both experiments are consistent overall, which validates the improved super-pulse method presented in the previous section. The slight differences between the two data sets are likely to be related to two factors (already explained above): (1) the elevation angle was not exactly the same in both experiments, and (2) the data presented in Fig. 6.3 corresponds to the actual transfer function (or intensity gain), whereas Fig. 7.3a shows the total intensity at the eardrum.

0° elevation. Results at 0° elevation for the unaltered-pinna condition are shown in left-most panel of Fig. 6.4. It is difficult to establish a correspondence...
between the notches observed at this elevation and those described at \(-45^\circ\) without having a continuous set of elevation-varying HRTF data. Results at 0° elevation (Fig. 6.4) can be only compared with those presented in Chapter 4 for the ear DB-061 (Fig. 4.6). From this comparison, it is possible to conclude that N1 is still present at 0° elevation (in Fig. 6.4) but its central frequency has increased from 7 kHz at \(-45^\circ\) to 8 kHz at 0° elevation for a source in the median plane (0° = 360° azimuth). It seems that N1b is also present at 0° elevation. Like N1, the centre frequency of N1b has increased from 8.3 kHz to 10.5 kHz for a source in the median plane. It is very difficult to say whether the spectral notches at higher frequencies correspond to either N2 or N3.

\(+45^\circ\) elevation. Results at this elevation are shown in Fig. 6.5. It is hard to establish a correspondence between the spectral features present at this elevation and those observed at \(-45^\circ\) and 0°. It should be noticed, however, that the higher the elevation angle, the higher the frequency at which spectral notches occur. The reason is that a second concha resonance at around 7.3 kHz appears at high elevation angles, [as found by Shaw (1975)]. This resonance peak effectively 'pushes' the spectral notches toward higher frequencies. As a result, the spectral shape at high elevations can be described (as in Chapters 4 and 5) as 'gradually decreasing toward the high-frequency end of the spectrum'.

6.3.2 Flange-filled condition

The HRTFs measured with the flange-filled are shown in the second panel (from the left of the picture) of Figs. 6.3, 6.4 and 6.5. By comparing the unaltered-pinna and the flange-filled conditions, it can be easily seen that the flange cavities clearly influence the presence of first spectral notch, whatever its frequency.

\(-45^\circ\) elevation. At \(-45^\circ\) elevation (Fig. 6.3), N1 and N1b disappear almost completely at azimuth angles far from the median plane when the flange cavities are filled. Even in the median plane (360° azimuth) the notches are barely distinguishable. In the lateral vertical plane, the notches N1 and N1b disappear but a new shallow notch appears between them at approximately 8.2 kHz. With regard to N2, it is still clearly observed in this set of results, which indicates that it is not associated to the flange cavities.
Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs

0° elevation. At 0° elevation (Fig. 6.4), N1 and N1b are substantially deeper when the flange cavities are filled than for the unaltered-pinna condition. Apart from this, there is no other major difference between the results obtained in the 'flange-filled' and the 'unaltered pinna' conditions.

+45° elevation. Results for this elevation are shown in Fig. 6.5. The first spectral depression (between 9.5 and 12.5 kHz) observed with the unaltered pinna disappears when the flange cavities are filled.

In summary, there are significant differences in the HRTFs obtained with the 'unaltered-pinna' and the 'flange-filled' ear-pieces, specially regarding the notch N1 (or notches, if we consider N1 and N1b). The way in which the notch is affected depends on the elevation of the sound source (as seen in Figs. 6.3, 6.4 and 6.5). This interesting result agrees with the observations made by Shaw (1982b) after filling the fossa of a standard KEMAR's ear with rubber.

However, it is not possible to conclude from this evidence that the effects observed when filling the flange cavities are actually a consequence of eliminating the 'acoustic resonances' possibly associated with those cavities. The resulting pinna shape after filling the flange cavities is altogether a different acoustic system (it is like a pinna with a flat flange). Therefore, it may be that the observed effects on N1 are related to the specific acoustic characteristics of the new 'flat-flange' system. For instance, it may be that reflections in the resulting flat flange (and not the elimination of the flange resonances) are responsible for the strong differences in the results for the 'flange-filled' and 'unaltered-pinna' conditions.

Some light can be shed into this issue by removing the pinna flange and measuring the HRTFs associated with the resulting 'no-flange' ear-piece. If the results are similar to those obtained with the 'flange-filled' ear-piece, then it can be concluded that the differences between the results obtained for the 'flange-filled' and the 'unaltered-pinna' ear-pieces are actually the result of occluding the flange cavities (i.e., of eliminating the possible resonances associated with those cavities). This suggestion is tested as follows.
6.3.3 No-flange condition

Results obtained with the 'no-flange' ear-piece are shown in the middle panel of Figs. 6.3, 6.4 and 6.5 (which correspond to \(-45^\circ\), \(0^\circ\) and \(+45^\circ\) elevation respectively). There are very few important differences below 12 kHz between these results and those obtained in the 'flange-filled' condition. The main differences are:

\(-45^\circ\) elevation. At this elevation (Fig. 6.3), the main difference is that the notch that appeared at 8.2 kHz in the 'flange-filled' condition for sources close to the lateral plane is at least 10 dB deeper when the flange is removed. N2 is still present in the HRTFs when the flange is removed, which indicates that N2 is not caused by the flange cavities or the flange itself.

\(0^\circ\) elevation. There are no substantial differences in the HRTF data obtained at this elevation with the 'no-flange' and the 'flange-filled' ear-pieces.

\(+45^\circ\) elevation. At \(+45^\circ\) elevation (Fig. 6.5), the spectral peaks appear more clearly resolved in the measurements made with no flange than in the 'flange-filled' data. There are clearly four spectral peaks, P1 at 2.8 kHz, P2 at 5 kHz, P3 at 7.3 kHz and P4 at 10.5 kHz. This result is in agreement with the observations made by Shaw (1975) with his mechanical reconstruction of the external ear. No spectral notches below around 12 kHz are observed.

In summary, there are only minor differences between the HRTFs obtained with the 'flange-filled' and the 'no-flange' ear-pieces. In other words, the HRTFs are similarly affected when the pinna flange is removed and when the flange cavities are occluded.

These results provide the necessary evidence to resolve the argument explained above with regard to the direct relationship between the notch N1 (and N1b) and the flange cavities (see previous section Flange-filled condition). Now, it can be said that the occlusion of the flange cavities affects largely the notches N1 and N1b. Moreover, the effect is different for different elevation angles (as explained above). In other words, the depth and the location dependency of N1 (and N1b) are largely determined by the openness of the flange cavities, rather than by the flange itself.
6.3.4 Concha without the crus helias

-45° elevation. Fig. 6.3 shows the results obtained with the concha from which the crus helias has been removed. There is virtually no difference below 13 kHz between these results and those obtained with the 'no-flange' ear-piece of the previous condition. Even N2 is still present, although its dependence on azimuth has changed slightly and has become very shallow at 360° azimuth. The azimuth dependency of the features does not follow any clear pattern above 13 kHz, but does not differ substantially from the observations made in the 'no-flange' condition.

0° elevation. Eliminating the crus helias (see Fig. 6.4) has a more noticeable effect at this elevation. The first notches (N1 and N1b) get narrower for sources close to the median plane (360°). Additionally, two well-defined local peaks appear at 11.5 and 12.8 kHz for sources close to the median plane. The notch at around 15 kHz gets very deep for sources at 350° azimuth.

+45° elevation. The main effect of removing the crus-helias (Fig. 6.5) from the concha is to shift the third peak from 7.3 kHz (see the 'no-flange' condition) to 8.2 kHz. Qualitatively, this result coincides with the observations by Shaw (1975) based on his simple-geometry metallic model of the external ear. Shaw found a spectral peak at 8.1 kHz in the data measured with the metal piece that simulates the crus helias in place. This peak shifted to 9.8 kHz when the measurements were made without the crus helias piece. A second noticeable effect is that the fourth peak at 10.5 kHz gets wider and more intense when the crus helias is removed.

6.3.5 Concha without the crus helias and without the tragus

Results obtained after removing the tragus from the previous ear-piece ('concha-no-crus-helias') are shown in the right-most panel of Figs. 6.3, 6.4 and 6.5, which correspond to -45°, 0° and +45° source elevation respectively.

-45° elevation. The removal of the tragus has a large effect on the results with respect to those obtained in the 'concha-no-crus-helias' condition. The 'depth' resonance of the concha P2 (see Chapter 3) decreases slightly in amplitude. This happens because the concha cavity becomes more exposed after the tragus is removed.
and, therefore, the resonance peaks become less sharp (Shaw, 1982b). The same is applicable to the spectral peak at around 10 kHz. A wide shallow spectral notch appears from around 6.5 kHz for a source in the median plane to 8 kHz in the lateral vertical plane (270° azimuth). For most azimuths this notch is no more than 5 dB deep. However, the notch disappears at 310° azimuth and is considerably deeper toward the lateral plane. Additionally, it is very interesting that N2 disappears for azimuth angles close to the median plane and becomes very narrow at azimuths close to the lateral plane. This indicates that N2 is associated with the tragus.

0° elevation. The notches N1 and N1b are largely affected by the removal of the tragus. An interpretation of the data is difficult, but it looks as though N1 and N1b disappear for azimuthal angles between 320° and 360° and a new local notch appears at 7.3 kHz for sources in the median plane. Spectral notches above 13 kHz also become shallower when the tragus is removed.

+45° elevation. At first sight, it seems that the removal of the flange has very little effect at this elevation at frequencies below 12 kHz. The main difference with respect to the 'concha-no-crus-helias' condition is that the peak at 8.2 kHz has moved to approximately 9 kHz. Above 12 kHz, the removal of the tragus produce spectral features that do not show a clear pattern of dependency on azimuth.

In summary, these results show that removing the tragus influences notably the HRTFs, particularly with regard to N2.
Figure 6.3. (Legend shown in the previous page.)
Figure 6.4. As in Fig. 6.3 but for a source at 0° elevation.
Figure 6.5: As in Figure 6.3, but for a source at +45° elevation.

Azimuthal angle (degrees)

Intensity (dB)

Frequency (kHz)

0.98
1.88
2.80
3.70
4.61
5.51
6.42
7.32
8.23
9.13
10.03
10.94
11.84
12.75
13.65
14.56
15.47
16.37

Unaltered pinna
Flange-filled
No-flange Concha-no-crus-helias
Concha-no-tragus

Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs.
Chapter 6: The contribution of individual anatomical features of the pinna to HRTFs

6.4 DISCUSSION

An important motivation for carrying out the experiments presented above was to find out whether individual anatomical features of the external ear generate specific spectral features observed in HRTFs or, on the contrary, they simply influence the spectral shape at a particular frequency range related, in some way, to the physical dimensions of the anatomical feature. Results presented above clearly show that altering or removing individual anatomical features influence the whole spectral shape rather than a limited frequency range. The same results indicate that, in some cases, individual spectral features are indeed influenced by the alteration or removal of individual anatomical features. For instance, it has been shown that filling the flange cavities or removing the flange have a very large effect in the first spectral notch (N1 and N1b), whatever its frequency. However, it would be wrong to infer from this result that the flange alone is responsible for N1 and N1b. No formal evidence of this has been provided. It can only be said that the flange has a strong influence on N1 and N1b, but it must be the interaction between the flange and the rest of the external ear that causes these two notches.

There are some cases, however, in which it is possible to say that a particular spectral feature is overall caused by an individual anatomical part. For example, the results at -45° elevation show that N2 is equally present, with approximately the same depth and azimuthal dependency, in the measurements made with the 'unaltered-pinna', the 'flange-filled' and the 'no-flange' conditions. In this case, it is possible to conclude that, in broad terms, N2 must be caused by the concha alone (that is, the full concha, with the crus helias and the tragus), at least at this elevation. Despite this evidence, a critical mind would argue that a closer look to the data shows the azimuthal dependency of N2 to be slightly different for the three conditions. This means that the flange must have an effect (however slight) on N2, even though the notch may be mainly produced by the concha.

These arguments lead to the conclusion that HRTFs can not be generally explained from the contribution of individual anatomical features, as is usually assumed. This is particularly true at frequencies above 6 kHz. (For further discussion see Shaw, 1975b). This is an important consideration to be taken into account when modelling...
the transfer function of the external ear. Therefore, it is inaccurate to look for an explanation to N1 solely based on the physical phenomena occurring in the concha, as proposed by Shaw and Teranishi (1968), Hebrank and Wright (1974) or in the previous chapter.

6.4.1 Compatibility of the results with the predictions of the diffraction/reflection model

The experimental data shown above indicate that N1 is only present when the flange cavities are unaltered. The generalisability of this result to other pinnae is supported by similar observations reported by Shaw (1982b) after filling the fossa of a standard KEMAR ear. Therefore, it can not be concluded that N1 is generated in the concha alone. This seems to contradict the model predictions at first sight, as the model predicts the existence of a notch similar to N1. However, the model results should only be compared with experimental data measured with a concha whose shape is similar to the approximated concha for which the model was evaluated (shown in Fig. 5.3). In this sense, the simulated HRTFs should only be compared (if at all comparable) with the 'concha-no-tragus' results as this is the ear-piece most similar to the geometry of the approximated concha for which the diffraction/reflection model was evaluated.

In the 'concha-no-tragus' data at -40° elevation (Fig. 6.3), a notch appears at approximately the same frequency as N1, moving from 7 kHz in the median plane to 8.3 kHz in the lateral plane. The notch is very shallow (around -5 dB; i.e., 5 dB below the 0 dB line) for sources close to the median plane (360°) but gets deeper (around -15 dB) at 280° azimuth. The diffraction/reflection model predicts a notch with these characteristics (see Fig. 5.5). Even the notch depth is approximately predicted by the model. It is important to remember, however, that this notch is not N1 (although it may be related), as N1 has an important contribution from the flange cavities. This observation has only been possible after the experiments carried out with and without the flange.

It is also observed that the overall spectral shape for the 'concha-no-tragus' condition is more similar to the predictions of the diffraction/reflection model than any one of the other four conditions tested. For instance, it has been shown that N2 is clearly present
at all azimuths in all the conditions tested apart from the 'concha-no-tragus' condition where it becomes very narrow between 270° and 310° and disappears at azimuths from 310° and 360°. This indicates that N2 is strongly related to the tragus and the crus helias, which were not included in the approximated concha for which the model was evaluated. Consequently, a much broader peak appears in the 'concha-no-tragus' data between 8 and 14 kHz for a source in the median plane. This peak coincides with that predicted by the diffraction/reflection model. Of course, it must be considered that the model results (Fig. 5.5) do not include the effect of the head and torso, which are also influence the spectral shape in the 'concha-no-tragus' condition.

The 'concha-no-tragus' data show a spectral depression at high frequencies (>14 kHz). It is hard to establish whether this depression corresponds to a single notch and, if so, its azimuthal dependency is not as clear as in the case of N1. It is possible that the notch N3 predicted by the diffraction/reflection model (Fig. 5.5) refers to this depression. However, unlike the experimental data, the model predicts a spectral notch with a clear azimuthal dependency. With regard to this discrepancy and others described above, it is important to bear in mind that even the 'concha-no-tragus' ear-piece used in the experiments has a different shape and dimensions and is much more intricate than the approximated geometry used in the model. This must have important effects along the whole spectrum but they will be particularly significant at high frequencies, where the wavelength is smaller or comparable to the concha dimensions.

6.4.2 Compatibility of the results with related research previously published

Previous related research was carried out by Teranshi and Shaw (1968), Shaw (1975, 1975b, 1982a). In these studies, individual anatomical features (ear-canal, concha cavity, crus helias, fossa and flange) were simulated by simple-geometric metallic pieces whose size and shape are only a crude approximation of the real features. The effect of individual anatomical features on HRTFs was then studied by gradually mounting these metallic pieces in order to build up a more sophisticated model of the external ear. The findings of these studies are in agreement with the results presented above whenever a direct comparison is possible (see Results section above).
Unfortunately, the methodology employed in those previous studies has important limitations which make a direct comparison very difficult. Firstly, the metallic pieces that simulate the anatomical features are only a simple approximation of the real system regarding their shape and texture. This prevents their 'mechanical model' of the external ear from simulating correctly spectral features above 10 kHz (Shaw, 197b), which are very sensitive to the shape of the concha and the position of the ear-canal. Additionally, those studies did not include the effect of the head and torso (see Chapter 3) as their mechanical pinna was not fitted into a manikin. In addition to this, their measurements with the metallic ear are untypical in the sense that they were made with the sound source very close to the mechanical ear — around 8 cm — (Teranishi and Shaw, 1968). These limitations are overcome in the experimental method adopted in this chapter. Of course, one should always bear in mind that using a KEMAR and pinnae replica is yet another approximation, although noticeably closer to the real system.

6.4.3 Psychophysical implications

As explained in the previous chapters, NI is believed to provide an important cue for the judgement of source elevation (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Bloom, 1977; Watkins, 1978). A different type of study by Gardner and Gardner (1973) demonstrated that localisation accuracy for sources in the anterior part of the median plane deteriorates as the pinna cavities are progressively occluded. They (Gardner and Gardner) showed that simply by occluding the fossa and the scapha simultaneously, the localisation error index increases to over 45%, that is more than 30% above its value in normal conditions (i.e., with unoccluded pinnae). Combining the information from these psychophysical studies, one would intuitively assume that occluding the flange cavities must have a noticeable influence on NI. The results presented above have shown that this is actually the case. In other words, the conclusions drawn here regarding the effect of the flange cavities are compatible with the psychophysical evidence.
6.5 SUMMARY

An improved method for measuring HRTFs has been presented and validated in this chapter.

The improved method has been used for measuring the effects of altering the anatomical structure of the pinna. The effects of filling the flange cavities, and gradually removing the flange, the crus helias and the tragus has been investigated and characterised.

The results obtained from this investigation indicate that:

(1) In general, it is not possible to establish a one-to-one correspondence between specific spectral notches observed in HRTFs and particular anatomical elements of the pinna. This means, for instance, that it is wrong to say that N1 is exclusively generated in the concha. In fact, data presented above indicate that the depth and location-dependency of N1 is largely determined by the flange cavities.

(2) Altering particular anatomical elements of the pinna affects the whole spectral shape and not only the frequency range whose corresponding wavelength is comparable to the physical dimensions of the anatomical element.

Therefore, these findings indicate that modelling the transfer function of the pinna by modelling and combining the transfer function of each individual anatomical element is not possible (see Chapter 2, Background). The acoustic interactions between different structural elements of the pinna largely influence the spectral shape at least for high frequencies (>7 kHz), and therefore, must be included in the formulation of a general model for the pinna.

Finally, with regard to the notch N2, data presented above indicate that its characteristics are largely (but not exclusively) determined by the tragus.
CHAPTER 7

A computer model of the auditory nerve response to HRTF stimuli using a linear gamma-tone auditory filter bank

7.1 INTRODUCTION

There is psychophysical evidence that pinna-based spectral features are used as localisation cues (Blauert, 1969/70; Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Bloom, 1977; Watkins, 1978; see Chapter 2, Background). This evidence indicates that the spectral characteristics of the stimulus must be signalled in some way to the central nervous system. If that is the case, the spectral information of the stimulus must be encoded in the discharge patterns of auditory nerve (AN) fibres.

However, there are certain properties of the transduction process in the peripheral auditory system which suggest that there are limits to the amount of spectral information that can be preserved in the AN response. For instance, it is known that AN fibres have a limited dynamic range (e.g., Sachs and Abbas, 1974; Winter et al., 1990). Therefore, one would expect the spectral features of the stimulus to be encoded in the auditory nerve response only at signal levels within the dynamic range of the fibres (i.e., above threshold and below saturation). Additionally, the mechanical filtering at the cochlea is very likely to distort in some way the original spectral shape of the stimulus (see below).

This chapter is concerned with investigating (1) to what extent the spectral features of the stimulus spectrum are observed in the discharge patterns of the AN fibres, and (2) what are (if any) the limiting factors. The investigation is done by modelling the
response of a population of AN fibres to stimuli containing the spectral characteristics of human head-related transfer functions (HRTFs).

The model employed (described below) is comprised of a number of stages that can be summarised as follows. An HRTF stimulus is generated by convolving either broadband noise or a Dirac pulse with the head-related impulse response associated with a particular source location. The resulting convolved signal is then filtered through a linear gamma-tone filter bank (Patterson et al., 1992) that simulates the basilar membrane filtering. The mechanical to neural transduction is simulated by using the inner hair-cell (IHC) model by Meddis (1986, 1988) (see also, Meddis et al., 1990) together with a spike-generation routine which converts the probability of firing to actual spikes and includes the refractory effects of the AN fibres.

It has been explained in previous chapters that location-dependent spectral features appear at frequencies above 5 kHz. This means that they occur above the cut-off of phase locking in AN fibres (Palmer and Russell, 1986). Therefore, the representation of spectral cues in the AN must be based on the discharge rate of the fibres alone. In other words, the temporal-place representations useful in speech analysis (Sachs et al., 1988) cannot contribute to the AN encoding of pinna-based spectral cues. For this reason, the analysis is done by presenting the output from the model in the form of rate profiles (Sachs and Young, 1979; Rice et al., 1995), in which the number of spikes or the probability of a spike occurrence are represented as a function of fibre’s centre frequency.

Some effort is given to obtain a set of parameters which simulate low, medium and high spontaneous-rate auditory nerve fibres (usually referred to as LSR, MSR and HSR fibres respectively). These parameters are then used to obtain rate profiles associated with each one of the three fibre types. The information contained in the simulated rate profiles is compared against the spectral information of the original HRTF. Because MSR and LSR fibres have a much wider dynamic range than HSR fibres, one would expect the spectral features of the stimulus to show much more clearly in the rate profiles of LSR and MSR fibres than in those of HSR fibres. Additionally, the rate threshold is higher for MSR and LSR fibres than for HSR fibres, which indicates that LSR and MSR fibres are likely to encode the spectral information at high stimulus levels. These hypotheses are tested below with the model and the results are discussed in the context of the experimental data available.
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

The model is also used to compare the quality of the rate representation of the stimulus spectrum at the stimulus onset and in the steady-state. It is known that AN nerve fibres have a much wider dynamic range at the stimulus onset than in the steady-state. Therefore, it is reasonable to hypothesise that the quality of the response will be better at the onset of the signal than in the steady-state. It has been previously reported that this is case (Delgutte and Kiang, 1984a,b). We are interested in finding out to what extent this behaviour can be reproduced by the simple linear model presented below.

It must be pointed out that the work presented in this chapter constitutes just a preliminary study in which the effects of some modelling parameters and the limitations of the proposed linear model are investigated and discussed. A much more sophisticated model which includes the nonlinear characteristics of the cochlear filtering is presented in the next chapter. That new, more comprehensive model is used in Chapter 9 to revisit the issues covered in the current chapter.

7.2 MODEL DESCRIPTION

The model is outlined in Fig. 7.1 and consists of the following stages: (1) stimulus generation; (2) read-in experimental head-related impulse response (HRIR); (3) HRTF-stimulus generation; (4) mechanical filtering by the basilar membrane; (5) mechanical to neural transduction at the inner hair-cell (IHC); and (6) spike-generation and refractory effects. Each one of those stages is explained individually below.

The model was implemented in LUTEar format (LUTEar, 1993) and, therefore, each model stage is available as a separate module.
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

Figure 7.1. Processing sequence of the model.
7.2.1 Stages (1), (2) and (3): HRTF-stimulus generation

The input signal to the model is the convolution of a stimulus signal, \( s(t) \), with a particular head-related impulse response, \( h(t) \). Two types of stimuli were used in the model evaluation (see below): broad-band noise and ideal Dirac-pulses or 'clicks'. The HRIR was obtained from the HRIR-database released by Gardner and Martin (1994, 1995) —see also Chapter 4—. We chose the HRIR measured at the left ear (model DB-061) of a KEMAR for a sound source at 0° azimuth, -20° elevation, placed at 1.4 m away from the ear. There was no particular reason for choosing that source position other than its corresponding HRTF (shown in Fig. 7.6a) presents an interesting number of spectral features (peaks and notches).

The convolution, \( u(t) \), of the stimulus signal with the HRIR was done using a discrete convolution of the form (adapted from Press et al., 1992):

\[
U_j = \sum_{k=1}^{M} H_k S_{j-k} \quad \text{where } 1 \leq j \leq N
\]  

(7.1)

where:

- \( U_j \) is \( j \)-th sample of the discrete convolved signal,
- \( H_k \) is the \( k \)-th sample of the discrete head-related impulse response,
- \( S_{j-k} \) is the \((j-k)\)-th sample of the discrete stimulus signal,
- \( M \) is the number of samples in \( H \),
- and \( N \) is the number of samples in \( S \).

In the implementation it is assumed that \( M < N \).

The sampling rate, \( f_s \), was determined by the sampling frequency of the provided HRIR which was 44.1 kHz (thus, \( dt = 1/f_s = 2.2675e-5 \text{ s} \)).

The resulting convolved signal has the spectral characteristics of the HRIR and was used as the input to the linear gamma-tone auditory filter bank.
7.2.2 Stage (4): Gamma-tone filter bank

The simulation of the bandpass filtering effect of the basilar membrane was achieved using the linear gamma-tone auditory filter bank available in LUTEar (LUTEar, 1993). The LUTEar version of this filter bank was adapted from the original version provided by Patterson et al. (1992). The gamma-tone filter characteristics (bandwidth and order) were developed from psychophysical studies and, therefore, the filter shapes do not represent physiological tuning curves. (For a review on the issue of correlation between psychophysical and physiological results see Pickles, 1988). The gamma-tone filter shape is a good approximation to the physiological tuning curves at the tip of the curves, but it fails to simulate the tuning-curve skirts at 30 dB below the best frequency of the filter.

It is important to stress that the gamma-tone filters are linear and, therefore, have the same bandpass parameters for all signal levels. This feature also deviates from the physiological results. It is now known that the filtering effect of the basilar membrane is nonlinear (Rhode, 1971; Robles et al., 1986; Rhode and Cooper, 1993). This nonlinear behaviour introduces a number of phenomena (e.g., two-tone suppression and distortion products) which must modify the spectral content of the original signal. The issue is further discussed below and the effects of including a nonlinear filter bank are examined in later chapters.

Despite the discrepancies between the gamma-tone filter bank characteristics and the physiological results, there is a degree of correlation between physiological and psychophysical results which justifies the use of this 'psychophysical' filter bank in our physiological modelling, at least as a first approximation.

The filter bandwidths were expressed as Equivalent Rectangular Bandwidths (ERB) (Glasberg and Moore, 1990). Thirty-three 4th-order filters were considered to simulate the filtering effect at 33 positions along the basilar membrane. The filters were centred at frequencies from 1 to 17 kHz, equally spaced every 500 Hz.
7.2.3 Stage (5): The mechanical to neural transduction

The output from each gamma-tone filter was passed to an inner hair-cell (IHC) simulator (illustrated in Fig. 7.2) which converted the mechanical motion of the basilar membrane at that point to the probability of a spike occurring in the post-synaptic auditory nerve. The IHC model and its properties are explained in detail in Meddis (1986, 1988) and Meddis et al. (1990).

It is assumed in the model that the probability of a spike occurrence is given by:

\[ p(t) = hc(t)dt \]  

(7.2)

where: 
- \( p(t) \) is the probability of a spike occurring during the period \( t \) to \( t+dt \),
- \( c(t) \) is the amount of transmitter in the cleft (i.e., between the IHC and its corresponding AN fibre),
- \( h \) is a parameter of the model and,
- \( dt \) is the sampling interval (unless otherwise stated \( dt = 2.2675e-5 \) s).

The model assumes that transmitter substance is released into the cleft from the IHC in amounts that depend on the permeability of the cell membrane, \( k(t) \). The membrane permeability, \( k(t) \), depends on the instantaneous signal amplitude, \( u(t) \):

\[ k(t) = \begin{cases} 
\frac{g[u(t)+A]dt}{u(t)+A+B} ; & \text{for } [u(t)+A] > 0 \\
0 ; & \text{for } [u(t)+A] \leq 0 
\end{cases} \]  

(7.3)

where: 
- \( u(t) \) is the instantaneous amplitude of the convolved signal after filtering,
- \( g, A \) and \( B \) are parameters of the model.
Figure 7.2. Diagram illustrating the operation of the (synaptic effects) hair cell model [Adapted from Meddis et al. (1990)].
In the period $t$ to $t+dt$, the amount of transmitter released into the cleft is given by $k(t)q(t)dt$, where $q(t)$ is the amount of transmitter in a free store lying close to the IHC membrane (Fig. 7.2). During the same time period, the re-uptake of transmitter from the cleft takes place at a rate $rc(t)$. However, some of the transmitter is lost completely from the system at a rate $lc(t)$. Therefore, the evolution of the cleft contents can be expressed as follows:

$$\frac{dc(t)}{dt} = k(t)q(t) - lc(t) - rc(t)$$  \hspace{1cm} (7.4)

The recovered transmitter is taken briefly to a reprocessing store [whose contents are denoted by $\omega(t)$] before being returned to the free store. The returning of the transmitter from the reprocessing store to the free store is done at a rate $x\omega(t)$. Therefore, the evolution of the contents of the reprocessing store can be expressed as follows:

$$\frac{d\omega(t)}{dt} = rc(t) - x\omega(t)$$  \hspace{1cm} (7.5)

The transmitter contents of the free store, $q(t)$, are, therefore, incremented by the recovered transmitter through the reprocessing store, which is done at a rate $x\omega(t)$. At the same time, the contents of the free store are decremented by the release of transmitter from the free store to the cleft, which is done at a rate $k(t)q(t)$. Apart from these two processes, the amount of transmitter in the free store is incremented by some new transmitter that is being manufactured at a transmitter 'factory'. The new transmitter is manufactured at a rate $y[M-q(t)]$. The new transmitter is manufactured to compensate the losses from the cleft. The evolution of the transmitter contents in the free store can be expressed as follows:

$$\frac{dq(t)}{dt} = y[M - q(t)] + x\omega(t) - k(t)q(t)$$  \hspace{1cm} (7.6)

Eqs. (7.4), (7.5) and (7.6), together with equation (7.3) describe the dynamics of the mechanical to neural transduction model by Meddis (1986, 1988). $x$, $l$, $r$, $y$ and $M$ are parameters of the model, together with $g$, $A$, $B$ of Eq. (7.3) and $h$ of Eq. (7.2).
The model explains a large number of AN phenomena (Meddis, 1986, 1988; Meddis et al., 1990; Hewitt and Meddis, 1991). However, we will concentrate mainly on its ability to simulate the rate-intensity (RI) curves of AN fibres, as the AN representation of pinna-cues must be done merely in terms of the fibre's discharge rate (see Chapter 2: Background). In this sense, the model is initially evaluated with the set of parameters for a HSR fibre given by Meddis et al. (1990). Then, alternative sets of parameters are presented (see section 7.3 below) with which the model simulates approximate RI curves for HSR, MSR and LSR fibres.

The model is implemented with only one hair cell (i.e., one AN fibre) per frequency channel. However, the model assumes that a large number of fibres are present within each channel and the probability of firing for one fibre is the same for all fibres of the same type. Additionally, the model is evaluated separately for different fibre types. When examining the rate profiles for each fibre type, we will assume that all fibres have the same RI curves (i.e., the same rate- and saturation-thresholds and the same dynamic range) across all frequency channels. These assumptions are made in the interest of computational speed. In the real case, however, fibres that are classified as the same type according to their spontaneous rate, may differ in other important characteristics (like their rate- and saturation-thresholds and their dynamic range) which must certainly influence the quality of the rate representation of the stimulus spectrum. The implications of these assumptions are discussed below.

7.2.4 Stage (6): Spike generation and refractory effects

The output from the inner hair-cell simulator is given as the probability of a spike occurrence in the post-synaptic AN afferent fibre. However, whether a spike actually occurs or not also depends on the recent history of the fibre. It is known that AN fibres have a refractory period of around 1 ms, which means that when a spike occurs in the post-synaptic fibre, the same fibre will not fire until 1 ms later. This refractory effect has been incorporated in the model. The way in which the probability of firing, \( p(t) \), is converted to spikes is as follows:

\[
spike(t) = \begin{cases} 
1, & \text{if } \left[ \left[ p(t) > R \right] \text{ and } \left[ spike(t - \tau) = 0 \right]\right] \\
0, & \text{otherwise}
\end{cases}
\]  

(7.7)
where: $\text{spike}(t)$ represents the spike function at instant $t$,

$R$ is a random number, $0 < R < 1$,

$\text{spike}(t-t)$ represents the spike function at instant $t-t$, with $0 < t < 0.001s$.

### 7.3 Parameters for the IHC Model: Rate-Intensity Curves

The parameters for the IHC model given in Meddis et al. (1990) were initially used for evaluating the model (see below). These parameters (shown in Table 7.1) simulate the rate-intensity (RI) curves for a HSR fibre and also other fibre characteristics like adaptation time-constants and synchronisation indices at 1 and 5 kHz. The onset and steady-state RI curves for these parameters are shown in Fig. 7.3. The RI curves were obtained for a centre frequency (CF) of 1 kHz. The stimulus was a 300-ms tone burst ($f = 1$ kHz) with 10 ms rise-time ($dt = 5e-5$ s). AN spikes were collected for a total of 250 noise bursts. The onset spike rate was calculated as the maximum spike rate over the first 10 ms of the signal. The steady-state spike rate was calculated by averaging over the last 150 ms of the signal.

\footnote{Notice that $A$ and $B$ differ from the values published in Meddis et al. (1990). This discrepancy is a consequence of using different intensity references in the model implementation. The values for $A$ and $B$ given in Table 7.1 generate identical results to those published in Meddis et al. (1990) when the signal intensity is expressed in dB SPL (as it is the case in LUTEar).}
Figure 7.3. (a) Onset rate-intensity curve simulated by the model using the set of parameters given in Meddis et al. (1990) for a HSR fibre — see Table 7.1 —. (b) Steady-state rate-intensity curve simulated by the model with the same set of parameters. N.B. different scales for y-axis.
Because of our interest in comparing the AN rate representation of HRTF stimuli for LSR, MSR and HSR fibres, alternative sets of parameters were obtained for simulating the RI curves of the three types of fibres. These parameters (shown in Table 7.1) were obtained from the parameters given in Meddis et al. (1990) by the laborious 'hill-climbing' approach of changing one parameter at a time and noting the effect. If the effect was beneficial, this set of parameters was used as a new starting point; otherwise, it was necessary to revert to the previous set and make a different change. The guidance given in Meddis et al. (1990) about the effects of changing each parameter was followed in the process.

Notice that the final parameter sets for the three fibre types only differ in the values of $A$ and $B$. This restriction was specifically imposed in the parameter optimisation process. It attempts to account for the physiological evidence that the three types of AN fibres differ only in the size of their synapses (Liberman et al., 1990), thus, in the permeability of the membrane, $k(t)$. This issue is fully addressed and discussed in the next chapter.

<table>
<thead>
<tr>
<th>Parameter description</th>
<th>Symbol</th>
<th>HSR, Meddis et al. (1990)</th>
<th>HSR (new)</th>
<th>MSR (new)</th>
<th>LSR (new)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permeability constant $A$ (1/s).</td>
<td>$A$</td>
<td>100</td>
<td>240</td>
<td>75</td>
<td>700</td>
</tr>
<tr>
<td>Permeability constant $B$ (1/s).</td>
<td>$B$</td>
<td>6000</td>
<td>5000</td>
<td>48000</td>
<td>6000000</td>
</tr>
<tr>
<td>Release rate (1/s).</td>
<td>$g$</td>
<td>2000</td>
<td>500</td>
<td>5000</td>
<td>5000000</td>
</tr>
<tr>
<td>Replenishment rate (1/s).</td>
<td>$y$</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
</tr>
<tr>
<td>Loss rate (1/s).</td>
<td>$l$</td>
<td>2580</td>
<td>1650</td>
<td>1650</td>
<td>1650</td>
</tr>
<tr>
<td>Reprocessing rate (1/s).</td>
<td>$x$</td>
<td>66.31</td>
<td>170</td>
<td>170</td>
<td>170</td>
</tr>
<tr>
<td>Recovery rate (1/s).</td>
<td>$r$</td>
<td>6580</td>
<td>8500</td>
<td>8500</td>
<td>8500</td>
</tr>
<tr>
<td>Max. no. of transmitter packets in free pool.</td>
<td>$M$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Firing rate (1/s).</td>
<td>$h$</td>
<td>500000</td>
<td>500000</td>
<td>500000</td>
<td>500000</td>
</tr>
</tbody>
</table>

Table 7.1. Parameters for the mechanical to neural transduction model (Meddis, 1986, 1988; Meddis et al., 1990).
Fig. 7.4 shows a comparison between experimental steady-state RI curves [adapted from figure 1 of Winter and Palmer (1991)] and the RI curves produced by the model with the new sets of parameters for LSR, MSR and HSR fibres. The CF of the fibre was 2.1 kHz. Spikes were collected over 250 runs of 300 ms tone-bursts ($f = 2.1$ kHz), with 2.5 ms rise-time ($dt = 5e-5$ s). The steady-state average spike rate was calculated by averaging over the last 150 ms of the signal. The simulated RI curves at the stimulus onset are also shown (Fig. 7.4b). The onset rate was calculated as the maximum rate over the first 10 ms of the stimulus.

There are discrepancies between the simulated and the experimental steady-state RI curves shown in Fig. 7.4. The most obvious is that the saturated rate of the simulated fibres is lower than the experimental value. However, this was not considered as a serious discrepancy with the physiological data for two reasons; firstly, because it is known that there is a large diversity of saturated rates amongst HSR fibres (Winter et al., 1990); and secondly, because the experimental RI curves (Winter and Palmer, 1991) were obtained by averaging over the whole duration of the tone burst (50 ms) and, therefore, their average experimental rates must have been affected by the high firing rates associated with the stimulus onset.

More important, however, are the discrepancies regarding the overall shape of the RI curves. The implemented model is unable to simulate the characteristic sloping saturation of MSR fibres (Sachs and Abbas, 1974; Winter et al., 1990). This characteristic increases the dynamic range of the fibre which is a crucial factor in determining the AN rate representation of the stimulus spectrum (see below). Additionally, the simulated fibres always saturate (in the case of LSR this occurs at intensities above 110 dB SPL and, therefore, it is not shown in the figure). This is not always the case in the measured RI curves (Sachs and Abbas, 1974; Winter et al., 1990). This factor also affects notably the AN rate representation of the stimulus spectrum (see below).

It is known that the sloping saturation of MSR fibres is a direct consequence of the nonlinear filtering of the cochlea (Yates et al., 1990; Yates, 1990). Therefore, a nonlinear model of the cochlear mechanical filtering is required to simulate this important characteristic. This issue is fully addressed in the next chapter where a nonlinear model is proposed and its predictions are tested against experimental RI curves of LSR, MSR and HSR fibres.
Figure 7.4. (a) Simulated steady-state rate-intensity curves. (b) Simulated onset rate-intensity curves. (c) Experimental rate-intensity curves [reproduced from Winter and Palmer (1991)]. CF = 2.1 kHz. The modelled curves were generated with the new sets of parameters for the mechanical-to-neural transduction model (Meddis, 1986, 1988) that are given in Table 7.1. Notice that the spontaneous rate of the experimental and modelled curves is approximately the same. However, the saturation rate is different (see text for details).
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

7.4 MODEL EVALUATION: AUDITORY-NERVE RATE PROFILES IN RESPONSE TO HRTF STIMULI

The proposed model was used for simulating the AN response to stimuli containing the spectral features of human HRTFs. The simulated AN response is presented as a rate profile, in which the firing rate of the AN fibres is represented as function of the fibres' CF.

The following issues were examined with the model: (1) the extent to which the spectral features of the stimulus are preserved in the output profiles from the linear gamma-tone filter bank; (2) whether the quality\(^2\) of the AN rate representation of stimulus spectrum is better at the stimulus onset than in the steady-state, particularly at high signal levels (i.e., the effects of fibre adaptation on the quality of the AN representation); (3) the quality of the AN rate profiles for LSR, MSR and HSR fibres at low, moderate and high signal levels.

The stimulus was generated by convolving either broad-band noise or a Dirac pulse with the HRIR corresponding to a sound source at 0° azimuth, -20° elevation (see section 7.2.1 above). The model evaluation procedure, the parameters employed and the results obtained in these studies are explained in the following sections.

7.4.1 Output profiles from the gamma-tone filter bank

Fig. 7.5 shows the output profile from the gamma-tone filter bank in response to the HRTF stimulus at three different signal levels: 30, 50 and 80 dB SPL (rms). The stimulus spectrum is shown at the bottom of the figure for comparison. In this particular case, the HRTF stimulus was generated by convolving the considered HRIR (see above) with a Dirac pulse.

\(^2\) The word 'quality' must be understood as the degree of similarity between the stimulus spectrum and its corresponding AN rate profile.
The results presented in Fig. 7.5 show that the fine spectral characteristics of the spectrum are lost in the AN response. Additionally, although the broad spectral characteristics of the stimulus are visible in the gamma-tone output profiles, the peak-to-notch amplitude is considerably smaller in the filter output profile than in the stimulus spectrum. The effect appears to be stronger for higher CFs. Therefore, the gamma-tone filter bank has the effect of 'smearing' the stimulus spectrum. This smearing effect is attributed to the energy spreading across CFs introduced by the filters' bandwidths. The overlapping of the gamma-tone filters causes the spectral energy at the peaks to spread over to other filters with close CFs. As a result, the spectral notches become shallower. Because the filter bandwidth increases as a function of CF (Patterson et al., 1992), the smearing effect is more obvious at high CFs.

Another important result from Fig. 7.5 is that the output profiles are identical at all three signal levels considered. As explained above (section 7.2.2, Gamma-tone filter bank), the characteristics (bandwidth and filter order) of the gamma-tone filters are the same for all signal levels, which explains why the output profiles are identical at 30, 50 and 80 dB SPL in Fig. 7.5 (see Discussion below).
Figure 7.5. Output from the linear gamma-tone filter bank when the stimulus is a Dirac pulse (click) convolved with the HRIR corresponding to a sound source at $-20^\circ$ elevation, $0^\circ$ azimuth. Three different stimulus intensities are represented 30, 50, and 80 dB SPL (rms). The bottom series represents the spectrum of the HRIR. N.B. the x-axis units must be understood as frequency (kHz) for the HRIR spectrum but centre frequency (kHz) for the gamma-tone output profiles.
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

7.4.2 Onset and steady-state rate profiles for HSR fibres in response to HRTF-filtered broad-band noise

The proposed model was used to investigate the differences in the quality of the rate profiles associated with the stimulus onset and the steady-state at various signal levels.

The IHC parameters for HSR fibres provided by Meddis et al., (1990) — specified in Table 7.1 — were used to generate post-stimulus time histograms (PSTHs) at the following signal levels: 35, 50, 85 and 110 dB SPL (rms). The PSTHs (bin-width = 1 ms) were produced by collecting the spikes generated by the model for a total of 250 HRTF-filtered Gaussian-noise segments of 50 ms of duration with a rise-time of 10 ms. Because the noise segments were of short (finite) duration, they could not possibly have a completely flat spectrum. As a result, there was the danger that the own spectral characteristics of the noise segments would be reflected in the simulated AN rate profiles, which could lead to misinterpreting the final results. In order to prevent this problem, a different noise sequence was used for each one of the 250 HRTF-filtered noise bursts, so that the random variations in the noise spectrum were 'cancelled' over the 250 runs.

In the model evaluation, the initial state of the various reservoirs in the IHC simulator was calculated assuming a long period of silence previous to stimulation (see Meddis et al., 1990). When calculating the PSTH, it was assumed that each one of the 250 noise bursts was preceded by an infinitely long period of silence [i.e., the amount of transmitter in the cleft, c(t), was reset to the value that it would have after a long silence period, Meddis et al. (1990)].

Onset-rate profiles were calculated from the simulated PSTHs by finding the maximum discharge rate over the first 10 ms of the signal. Steady-state-rate profiles were calculated by averaging over the last 20 ms of the signal duration. Additionally, average-rate profiles were calculated by calculating the average discharge rate over the full 50 ms of the signal duration.
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

Results: onset and steady-state rate profiles for HSR fibres

Fig. 7.6 shows the resultant AN rate profiles compared with the considered HRTF (Fig. 7.6a). Here is the analysis of the results.

• **35 dB SPL.** At 35 dB SPL, the spectral characteristics of the stimulus are just visible in the steady-state and the average rate profiles (Figs. 7.6c and 7.6d). However, they are not visible in the onset rate profiles (Fig. 7.6a). An explanation to this result can be attempted from looking at the RI curves for the considered HSR fibres (Fig. 7.3a). However, it must be first pointed out that such RI curves were obtained for pure-tone stimuli and, therefore, its corresponding rate threshold is slightly lower than the expected for broad-band stimuli (a full explanation to this fact is given in Chapter 9, section 9.3.1 Rate-intensity curves for broad band noise). Bearing this in mind, and considering that the rate threshold of the HSR fibre appears to be approximately 10 dB lower in the steady-state RI curve that in the onset one (Fig. 7.3a), it is very likely that a 35 dB SPL signal is just below threshold at the onset, but just above threshold in the steady-state response.

• **50 dB SPL.** At this intensity, the broad spectral characteristics of the stimulus are clearly visible in the three different types of rate profiles considered (i.e., onset, steady-state and average). This must occur because a 50 dB stimulus is within the dynamic range of the AN fibres (see Fig. 7.3). However, the AN rate profile appears to be a smeared version of the stimulus spectrum. This is, in part, a consequence of smearing effect by the gamma-tone filters presented above. It is very clear, therefore, that the quality of the AN rate representation of the stimulus spectrum is largely determined by the output from the basilar membrane (BM) filtering.

However, in some cases the smearing of the AN rate profile may be accentuated by the limited dynamic range of the fibres (particularly in the steady-state response) which will flatten the AN rate representation of the stimulus spectrum at those features with intensities below (in the case of notches) or above (in the case of peaks) the rate- and saturation-thresholds respectively. This effect will be less severe in the onset rate profiles, as the onset RI curve shows a much wider dynamic range.

153
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

- **85 dB SPL.** At this intensity, the steady-state and average rate profiles (Figs. 7.6c and 7.6d) are 'flat' and do not show any of the spectral characteristics of the stimulus. This occurs because at this moderately-high intensity the fibres are already saturated (as shown in Fig. 7.3b). Notice that according to the steady-state rate profiles of Fig. 7.6c, the saturation rate of the fibres is around 120 spikes/s, whereas it is around 90 spikes/s in the RI curve for the same fibres (Fig. 7.3b). This apparent discrepancy is related to the fact that the steady-state rate in Fig. 7.3b was calculated as the average over the last 150 ms of a 300-ms stimulus, whereas that in Fig. 7.6c was calculated over the last 20 ms of a 50-ms stimulus. This indicates that the fibre is in fact not fully adapted at the last 20 ms of the 50 ms noise burst, and that longer stimuli should be used if the actual steady-state discharge rate was to be studied.

At 85 dB SPL, however, *some* of the broad spectral characteristics (P1 and P4 for instance) are still visible in the onset rate profiles (Fig. 7.6b). This must occur because the onset response of the fibres is not yet saturated at this intensity, as shown in Fig. 7.6b. These results indicate, therefore, that at moderately-high signal levels some spectral characteristics of the stimulus may be indeed signalled to higher nuclei of the auditory system only in the onset response (or onset rate profiles) of AN fibres.

- **110 dB SPL.** At this intensity, the spectral characteristics of the stimulus are not observed even in the onset rate profiles. This is somewhat surprising since the onset RI curve of the considered fibre (Fig. 7.3a) is not saturated around 110 dB.

In summary, the steady-state AN rate response encodes the spectral characteristics of the stimulus at low and moderate signal levels (35 and 50 dB SPL rms), but fails to do so for moderately high signal levels (80 dB SPL rms). In this case, results presented above indicate that the spectral characteristics of the stimulus are still encoded in the onset response. At even higher signal levels (110 dB SPL rms), the spectral information is simply not encoded at all in any of the investigated rate profiles.
Figure 7.6. (a) Spectrum of the HRIR corresponding to a sound source at 0° azimuth, and -20° elevation. (b) Onset rate profile at four signal levels: 35, 50, 85 and 110 dB SPL (rms). The onset rate for each CF represents the maximum rate over the first 10 ms of the stimulus duration (50 ms). (c) Steady-state rate profiles for the same stimulus and intensities as in b). The steady-state rate for each CF represents the average over the last 20 ms of the 50 ms noise burst. (d) Average rate over the full duration of the noise burst (50 ms). N.B. different y-axis scales.
7.4.3 Rate profiles for LSR, MSR and HSR fibres at different stimulus intensities

In this study, 'clicks' (Dirac pulses) were used as the stimulus. Therefore, the input to the model was a click convolved with the considered HRIR (see above). In this case, the last stage of the model sequence (see Fig. 7.1) was not included in the model evaluation. This means that the model output was given directly from the IHC simulator as probabilities of spike occurrences (rather than actual spikes) and that the refractory effects in the post-synaptic fibre were not taken into account. This approximation was made in the interest of computation speed, as only one model run was necessary\(^3\) instead of the 250 runs employed to calculate the PSTH in the first study. For each frequency channel, the output from the model was expressed as the total probability of firing during a 1 ms window (or bin-width). This value was calculated by summing up the probability of firing at each one of the sampling intervals \((dt = 2.26e^{-5} \text{ s})\) contained within the 1 ms window. The data collected were then analysed to obtain the maximum probability of firing across all bins. An approximate spike rate (in spikes/s) was then obtained by dividing the maximum probabilities by the window duration (or bin-width), i.e., by dividing by \(1.0e^{-3} \text{ s}\).

In this study, the model was evaluated using the three new sets of parameters (given in Table 7.1) that simulate LSR, MSR and HSR fibres. For each fibre type, the model was evaluated at three different signal levels: 30, 50 and 80 dB SPL (rms). The results obtained are shown as follows.

**Results: Rate profiles for LSR, MSR and HSR fibres**

The results (shown in Fig. 7.7) represent the maximum discharge rate (calculated as explained above) as a function of the fibres' CF. It must be stressed that the discharge rates presented below are approximate. Because a click was used as the stimulus, the

---

\(^3\) Notice that because a click —thus, flat spectrum— was used as the source stimulus, there was no danger of the AN rate profiles being affected by the own spectral content of the source, as it happens when short noise bursts are used (see above, section 7.4.2). For this reason only one model run is enough if the output from the model is expressed as probabilities of spike occurrences, rather than actual spikes. Although this procedure only produces approximate results, it reduces the computational time enormously.
results may be interpreted as a kind of onset response. The results show the following features at the signal levels considered:

- **30 dB signal level.** Fig. 7.7a shows the results for a 30 dB SPL stimulus. It can be seen that the LSR rate profile shows no spectral features. According to Fig. 7.4, this happens because the rate-threshold of the LSR fibre is at around 45 dB SPL, which is above the intensity of the delivered signal. On the other hand, the rate profiles of both HSR and MSR fibres show a smooth version of the spectral features observed in the stimulus spectrum. They certainly show the spectral features present at the filter output (Fig. 7.5). It could be concluded that at this signal level, and for the simulated fibres, HSR fibres are more appropriate for encoding spectral information present in HRTFs.

- **50 dB signal level.** Fig. 7.7b shows the results for a 50 dB SPL stimulus. At this stimulus intensity, the spectral information of the stimulus is encoded in the rate response of both HSR and MSR fibres. The rate profile of LSR fibres still does not show any significant spectral feature, as expected from looking at the onset RI curve shown in Fig. 7.4c.

- **80 dB signal level.** The results at this signal level are shown in Fig. 7.7c. At this signal intensity, the spectral information seems to be encoded mainly in the rate response of MSR fibres. LSR fibres just begin to respond to the spectral variations of the stimulus. HSR fibres seem to be completely saturated at this signal level. This latter result is surprising considering that the represented rate profiles correspond to the maximum rate, thus they are a kind of onset-rate profiles. However, it can be explained by looking at the onset RI curve for the HSR fibre that is shown in Fig. 7.4c. It can be seen in the figure that the onset RI curve does not increase monotonically; it presents an 'inflection' point precisely around 80 dB SPL where the rate decreases locally from 80 to 100 dB before starting to increase again.

In summary, the results presented above show that MSR fibres are able to encode the spectral information of the stimulus at 30, 50 and 80 dB SPL. Of course, this is the result of them having a low rate-threshold and a wide dynamic range. HSR fibres, however, can only encode the information at low and moderate signal levels (30 and 50 dB SPL), as result of having a low rate-threshold and a short dynamic range. As it
was expected, LSR encode the spectral information at high signal levels (80 dB SPL) as a result of them having a very high rate-threshold and a wide dynamic range.

An attempt has been made to explain these results by examining the onset RI curves (Fig. 7.4c) of the LSR, MSR and HSR fibres used in the simulation. It should be noticed, however, that the discharge rates shown in Fig. 7.7 are higher than those in Fig. 7.4c for the same stimulus intensity. This result is more obvious at 80 dB SPL. This apparent inconsistency of the results is a direct consequence of not considering the refractory effects in Fig. 7.7, which reduce the number of spikes mainly at high intensities. It must be noticed, however, that the qualitative conclusions drawn above after the analysis of the rate profiles for LSR, MSR and HSR fibres are unaffected by the fact that the discharge rates of those rate profiles are only approximate.
Figure 7.7 Approximate rate profiles for the three fibre types, LSR, MSR and HSR, in response to a click with the spectrum of the HRIR shown in Fig. 7.6a. The maximum discharge rate is shown in the y-axis (see main text for details). (a) 30 dB SPL signal level. (b) 50 dB SPL signal level. (c) 80 dB SPL signal level.
Chapter 7: Modelling the AN response to HRTF stimuli using linear AN filter bank

7.5 DISCUSSION and CONCLUSIONS

A computer model of the auditory periphery for simulating the activity of LSR, MSR and HSR AN fibres has been presented. The model has been used to simulate the auditory-nerve response to HRTF-filtered stimuli. The model results indicate that AN rate profiles show the spectral characteristics of the stimulus only at signal levels within the dynamic range of the fibres. Additionally, the model suggests that at moderately-high signal levels, the quality of the rate representation at the stimulus onset is better than in the steady-state. This result has been attributed to the fact that the onset response presents a wider dynamic range (Fig. 7.3). The model results also suggest that the spectral features of the stimulus are mainly encoded in the rate profiles of HSR and MSR fibres for low signal levels, MSR fibres for moderate signal levels and MSR and LSR fibres for high signal intensities. These results have been attributed to the important differences in rate thresholds and dynamic ranges of LSR, MSR and HSR fibres.

7.5.1 Model limitations

An important result from this investigation is that the quality of the AN rate representation of the stimulus spectrum is largely affected by the mechanical filtering effect by the basilar membrane. In our case, a linear gamma-tone filter bank has been used to simulate such effect. It has been shown that this linear filter effectively smooths the spectrum, thus, reducing the peak-to-notch amplitude. It effectively reduces the spectral 'contrast'.

However, the implemented gamma-tone filter is only a linear approximation which fails to simulate the nonlinear characteristics introduced by the basilar membrane (Rhode, 1971; Cooper and Rhode, 1995; Robles et al., 1986). These nonlinearities cause important phenomena like two-tone suppression (Rhode and Cooper, 1993) and distortion products (Robles et al., 1993; Rhode and Cooper, 1993) which have been extensively reported at the AN (Arthur et al., 1971) level but are now known to occur also at the cochlea (Robles et al., 1993, Rhode and Cooper, 1993). These two phenomena (a good revision of the topic is given by Pickles, 1988) are important from our point of view because they must distort the AN rate representation of the stimulus spectrum when broad-band signals are used to stimulate the system (Yates et al., 1988).
1990). Two-tone suppression phenomena possibly have the effect of increasing the spectral contrast, whereas the generation of distortion products is likely to smooth the spectral shape and, therefore, the AN rate representation. Additionally, it is known that the cochlear nonlinearity determines the shape of the AN RI functions (Yates et al., 1990). This must also influence the quality of the AN rate representation of the stimulus spectrum because, as we have seen above, it is directly related to the shape of the RI curves. These significant phenomena have not been included in the computer model proposed above. However, they will be extensively examined in the next chapters.

7.5.2 Diversity of AN fibres: Implications in modelling the AN response to HRTF stimuli

In modelling the AN rate response to HRIR stimuli we have considered that all fibres of the same type have the identical RI curves across CFs, i.e., the same rate- and saturation-thresholds and the same dynamic range. It is known, however, that indeed there are fibres which are included in the same category of LSR, MSR and HSR according to their spontaneous rate\(^4\), but have, in fact, different rate- and saturation-thresholds, saturation rates and dynamic ranges (Liberman, 1988; Winter et al., 1990). This diversity has not been included in the model. A direct consequence of this approximation is that it is always possible to compare directly the simulated rate profiles with the stimulus spectrum. In experimental studies this is not the case (Rice et al., 1995). Even when the AN rate representation is examined separately for LSR, MSR and HSR fibres at different intensities, the corresponding rate profiles hardly resemble the spectral shape. This is very possibly related to the variability of characteristics among fibres of the same type which can not be controlled in experimental studies. In order to facilitate the comparison between experimental AN rate profiles and the stimulus spectrum it is common to normalise\(^5\) the measured discharge-rates according to the rate characteristics of each fibre. This procedure is not

\(^4\) Originally, the classification of AN fibres according to their spontaneous rate was done by Liberman (1978).

\(^5\) The normalisation is done as follows: \( NR = (R - SR)/(Sat - SR) \), where \( NR \) is the normalised rate, \( R \) is the measured rate for the fibre, \( SR \) is the spontaneous rate for the fibre, and \( Sat \) is the saturation rate for the fibre. Notice that this normalisation only accounts for differences in \( SR \) and \( Sat \) rate, but not for the differences in the dynamic range of the fibres. The issue is fully discussed in Chapter 9.
necessary for the model results as the normalisation is done 'by default'. This issue will be fully addressed in Chapter 9.

### 7.5.3 Psychophysical implications for sound localisation

The preliminary results presented above suggest that broad pinna-based spectral cues may be encoded to some extent in the AN rate response at signal levels up to around 80 dB SPL. Therefore, it is tempting to say that it is possible that the spectral features important for sound localisation, like the elevation-dependent spectral notch (N1) described in the previous chapters, are actually transmitted along the auditory brainstem. However, this conclusion is only preliminary as the results presented above are inconclusive in the sense that the important nonlinear effects of the cochlear filtering have not been considered. The issue is re-examined in Chapter 9 where a more sophisticated computer model of the peripheral auditory system that includes the nonlinear characteristics of the cochlear filtering is used for investigating the AN response to HRTF-stimuli in a more precise manner.
CHAPTER 8

A computer model of low, medium and high spontaneous rate auditory-nerve fibres using a dual-resonance nonlinear auditory filter bank

8.1 INTRODUCTION

A computer model of low, medium and high spontaneous rate (LSR, MSR and HSR respectively) fibres is presented and evaluated in this chapter. The model (shown in Fig. 8.1) incorporates the dual-resonance non-linear (DRNL) auditory filter bank developed by O'Mard and Meddis (1996). The incorporation of a nonlinear filter bank constitutes a substantial improvement with respect to the linear model proposed in the previous chapter for reasons which are given below. Note that the model presented below is developed with a view to using it for simulating the auditory nerve (AN) rate response to head-related transfer function (HRTF) stimuli (see Chapter 9). In this chapter, however, the model response is only tested for pure tone stimuli.

It was shown in Chapter 7 that the AN rate representation of the stimulus spectrum is closely linked to the shape of AN rate intensity (RI) curves. A model was proposed that reproduces AN RI curves using a linear gamma-tone filter bank to simulate the basilar membrane (BM) mechanical filtering effect. However, it is known that the BM filtering is nonlinear (Rhode, 1971). The relationship between the BM oscillation and the stimulus intensity is linear for stimulus frequencies below the best frequency (BF) of the measurement site along the BM. For stimulus frequencies at and above BF, the relationship is only linear up to a threshold intensity (known as the 'compression threshold'). For stimulus intensities above this threshold, the growth is very slow.
with a slope of approximately 0.2. For stimulus intensities around 30 to 40 dB above the compression threshold, the relation becomes linear again.

It is known that this nonlinearity determines the shape of the AN RI curves for stimulus frequencies at and off the fibre's BF (Yates et al., 1990; Yates, 1990; Müller and Robertson, 1991; Cooper and Yates, 1994). For instance, the RI curve of MSR fibres for stimulus frequencies at and above BF, show a characteristic sloping saturation (Sachs and Abbas, 1974; Winter et al., 1990; Yates et al., 1990; Liberman, 1988). That is, the RI curve of these fibres do not show a growth from spontaneous to saturated rate with constant slope but, on the contrary, there is an intensity threshold above which the slope of the RI curve decreases. Therefore, saturation is reached only at much higher intensities. This sloping saturation is thought to be a direct result of the nonlinear relationship between BM displacement and stimulus intensity, which occurs within the dynamic range of the AN fibre (Sachs and Abbas, 1974; Winter et al., 1990; Yates et al., 1990).

A direct consequence of the cochlear nonlinearity is, therefore, an increase in the dynamic range of MSR fibres for stimulus frequencies at, and below BF. It was concluded in Chapter 7 that having a wider dynamic range improves the quality of the AN rate representation of the stimulus spectrum over a wider range of stimulus intensities. This important consideration cannot be simulated using a linear model of the basilar membrane filtering, which constitutes one of the arguments in favour of using a nonlinear filter bank.

In addition to influencing the shape of AN RI curves, the BM nonlinearity generates other phenomena which should be also included in our simulation because they must influence the AN rate representation of the stimulus spectrum. These phenomena are the presence of distortion products and two-tone suppression (see Pickles, 1988 for a complete review on the topic). Both phenomena were first observed in the AN response to complex stimuli. However, it is now known that they also occur at cochlea and are a direct consequence of the nonlinear filtering by the BM (Robles et al., 1993).

In a linear system, the output signal contains only the same frequencies as the input. However, in a nonlinear system, single tones will also produce harmonics. When two tones of close frequencies are input to a nonlinear system, the output will produce, in
addition, combination tones. That is, new tones whose frequency depends on the frequency of both input tones. Consequently, when the input to a nonlinear system is a broad-band stimulus (as in the case of HRTF stimuli), the spectrum of the output signal will depend on the relative amplitude of adjacent frequency components in the input signal. In other words, the spectrum of the broad-band input signal is modified after filtering through a nonlinear system. This is an important phenomenon which should be taken into account when simulating the AN rate representation of HRTF stimuli. It can only be simulated using a nonlinear filter bank, which is another reason in support of incorporating a nonlinear filter bank to our model.

The phenomenon of two-tone suppression occurs when two pure tones of close frequencies are presented simultaneously. If the frequency of tones are of certain critical values, the presence of one tone may suppress the system's response to the other tone alone. Two-tone suppression must also occur when the stimulus is a broad-band signal (as in the case of HRTF stimuli), in which case the spectrum of the output signal will be different from the spectrum of the input signal. Therefore, this important phenomenon should also be taken into account when simulating the AN rate representation of the HRTF stimuli. Two-tone suppression is believed to be a consequence of the nonlinear filtering of the basilar membrane (see Pickles, 1988, p. 152 for a revision on the topic), which is another reason that justifies the incorporation of a nonlinear model of the BM to our model.

The DRNL model of BM filtering developed by O'Mard and Meddis (1996) simulates all these phenomena. This is the main reason it has been included in the computer model of the peripheral auditory system presented below. Another important property of this model is that its parameters have been fitted to reproduce BM physiological data.

The incorporation of the DRNL filter bank is perhaps the main characteristic that makes the model presented below considerably more realistic than the model presented in the previous chapter. Additionally, an effort has been made to reflect in the new model all the physiological stages that are thought to contribute to the AN response in the real system. For instance, the model proposed below includes three new stages with respect to the model presented in the previous chapter: the bandpass filtering of the middle-ear, a model which relates the deflection of inner hair-cell (IHC) stereocilia with BM displacement (Shamma et al., 1986), and a computer model of the IHC
receptor potential (Shamma et al., 1986). Moreover, although the synaptic effects are simulated using the model by Meddis (1986, 1988), the flow of transmitter amongst the different reservoirs is described in terms of discrete release of individual transmitter quanta on a stochastic basis.

Another important new feature of the model proposed below is that it incorporates an exponential function to describe the release of transmitter at the IHC/AN synapse. The choice of an exponential transmitter release function is based on physiological evidence as explained below (Stevens, 1968; Sewell, 1984). This function proves to be very appropriate for simulating the exponential dependence of the AN fibres' spontaneous rate with the endocochlear potential, as found by Sewell (1984) (see below). An exponential function is also very convenient because it allows us to speculate on the physiological differences amongst LSR, MSR and HSR AN fibres in terms of their corresponding synaptic size.

An appropriate set of parameters is obtained at each stage of the model to simulate corresponding physiological data. Finally, the model is used to simulate the RI curves of LSR, MSR and HSR auditory-nerve fibres for stimulus frequencies at- and off-BF. The study is restricted to BFs above 4 kHz, since this is the frequency range at which HRTF location-dependent spectral cues are encoded in the AN rate response.

### 8.2 MODEL DESCRIPTION

The processing sequence of the model is shown in Fig. 8.1. It consists of the following stages: (1) stimulus generation; (2) outer- and middle-ear combined band-pass filter; (3) nonlinear mechanical filtering of the basilar membrane; (4) BM displacement to IHC stereocilia displacement coupling; (5) IHC cilia displacement to intracellular voltage transduction; (6) IHC intracellular voltage to transmitter release at the IHC/AN synapse; (7) Pre-synaptic effects: transmitter contents in the synaptic cleft as a function of the transmitter release; (8) Spike generation and refractory effects at post-synaptic AN fibres. Each one of these stages is described below in detail.
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

Figure 8.1. Processing sequence of the model.
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

The model was implemented using the LUTEar library (LUTEar, 1993). Each one of the model stages was implemented as a separate module in LUTEar with its corresponding set of parameters.

8.2.1 Stage (1): Stimulus generation

Unless otherwise stated, only pure tones are used as stimuli in the model evaluations presented below. The input to the model is a discrete version of a continuous sinusoidal signal, which is given by:

\[ S_i = A_p \sin(2\pi ft) \]  

where:  
\( S_i \) is the i-th element of the discrete signal,  
\( A_p \) is the peak amplitude of the pure tone,  
\( f \) is the frequency of the tone,  
\( dt \) is the sampling interval \( (dt = 1/\tau) \) and,  
\( i \) is an integer number, \( 0 < i < N \), with \( N \) being the number of samples. \( N \) is calculated from the stimulus duration \( (D) \) as follows: \( N = D/dt \).

The peak amplitude, \( A_p \), is calculated from the specified tone intensity, \( I \), operating with the expressions:

\[ I(\text{dB SPL}) = 20\log_{10} \left( \frac{A_{\text{rms}}}{A_{\text{ref}}} \right) \]  

and:

\[ A_{\text{rms}} = \frac{A_p}{\sqrt{2}} \]  

where:  
\( A_{\text{rms}} \) is the root mean square pressure amplitude,  
\( A_{\text{ref}} \) is the reference pressure which has been taken as 20 \( \mu \text{Pa} \),  
\( A_p \) is the peak pressure amplitude, and  
\( I \) is the pure tone intensity expressed in dB SPL.

The parameters for generating the pure tone stimuli are, therefore, the frequency, \( f \) (in Hz), the stimulus duration, \( D \) (in s), the intensity, \( I \) (in dB SPL), and the sampling interval, \( dt \) (in s). The resulting discrete signal is a series of numbers that describe the instantaneous pressure amplitude in \( \mu \text{Pa} \).
8.2.2 Stage (2): Middle-ear filtering

In physiological experiments, sound stimuli are usually delivered at the entrance of the animal's ear-canal. The middle ear, which transforms the sound pressure variations at the ear-drum into a sound pressure variation in the scala vestibuli, has a band-pass transfer function (Nedzelnitsky, 1980), with a peak at around 1 kHz. In our model, the transfer function of the middle-ear has been simulated with a band-pass filter (referred to as the pre-emphasis filter) as suggested by O'Mard and Meddis (1996).

The pre-emphasis filter is defined by the lower cut-off frequency \( f_{lc} = 1 \text{ kHz} \), the upper cut-off frequency \( f_{uc} = 30 \text{ kHz} \), and the filter order \( p = 2 \) (O'Mard and Meddis, 1996). These parameters were optimised to simulate the average audiometric threshold of the albino guinea pig (see Fig. 8.2). Additionally, an attenuation factor, \( \alpha_{PE} \), is included which accounts for the fact that, in some experimental data (see below), the threshold values of basilar membrane data are different from 0 dB SPL, even at the most sensitive BFs.

Fig. 8.2a shows the shape of the pre-emphasis filter. Fig. 8.2b shows the model middle-ear threshold compared with psychophysical audiograms of guinea pig and chinchilla.

The output from the pre-emphasis filter is provided as the input to the DRNL filter, which is described in the next section. Notice that the DRNL filter was designed (O'Mard and Meddis, 1996) to work together with the pre-emphasis filter described above.
Figure 8.2. (a) The shape of the pre-emphasis filter ($\alpha_{PE} = 0$ dB). (b) Model middle-ear threshold tuning curve compared with the psychophysical audiograms for the guinea pig (Prosen et al., 1978) and chinchilla (Clark and Bohne, 1986).
8.2.3 Stage (3): Dual-resonance nonlinear (DRNL) auditory filter bank

The pressure variations in the scala vestibuli are transformed into BM oscillations. However, the relationship between BM oscillations and stimulus intensity is nonlinear (see introduction). The dual-resonance nonlinear (DRNL) model of the BM filtering developed by O'Mard and Meddis (1996) simulates the nonlinear characteristics of the system. As the model is explained in detail elsewhere (O'Mard and Meddis, 1996), only a brief description is given here below. Additionally, the adequacy of this model is briefly discussed with regard to using it for simulating the AN response to HRTF stimuli, which will be done in the next chapter.

DRNL model overview

A schematic representation of the DRNL filter is shown in Fig. 8.3. The output signal is the sum of the outputs from two different paths: the wide-filter path, and the narrow-filter path. In the wide-filter path, the input signal is filtered through a wide, linear, 4-th order gamma-tone filter (Patterson et al., 1992) with a centre frequency \( \omega_c^W \) and a bandwidth, \( \beta^W \). The signal is then filtered through a low-pass filter (with a cut-off frequency \( \phi \omega_c^W \)), and attenuated with an attenuation \( \alpha^W \). In the narrow-filter path, the input signal is filtered through a narrow, linear, 2-nd order gamma-tone filter with a centre frequency \( \omega_c^N \) and a bandwidth \( \beta^N \). The output of this narrow gamma-tone filter is then subject to a nonlinearity of the form:

\[
R(t) = \begin{cases} 
    r(t) ; & \text{if } |r(t)| \leq \Theta_{Th} \\
    \text{sign}[r(t)]\left[\Theta_{Th} + |r(t)| - \Theta_{Th}\right]^\nu ; & \text{if } |r(t)| > \Theta_{Th}
\end{cases}
\]  

(8.3)

where: \( r(t) \) is the input signal after filtering through the narrow gamma-tone filter, \( \Theta_{Th} \) is the threshold amplitude below which no compression is applied and, \( \nu \) is the compression exponent.
The input/output (I/O) curve for this function is linear for input amplitudes below the compression threshold, $\Theta_{Th}$. For input amplitudes above this threshold, the signal is compressed as indicated in Eq. (8.3).

The output from the nonlinearity is filtered again through a second linear gamma-tone filter identical to the first one in this narrow path.

The main parameters of the model are, therefore, $\omega_{cf}^N$, $\beta^W$, $\alpha^W$, $\phi$, $\omega_{cf}^N$, $\beta^N$, $\Theta_{Th}$ and $\nu$. Because of the limitations in the current experimental techniques, BM experimental data are only available at three BFs: approximately 300 Hz (Cooper and Rhode, 1995), 8 kHz (Sellick et al., 1982), and 18 kHz (Robles et al., 1986). O'Mard and Meddis (1996) provide an optimised set of parameters for their DRNL model to simulate the BM data at those three BFs. These parameters are shown in Table 8.1.

<table>
<thead>
<tr>
<th>Narrow Filters</th>
<th>Wide Filters</th>
<th>Compression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{cf}^N$ (Hz)</td>
<td>$\beta^N$ (Hz)</td>
<td>$\omega_{cf}^W$ (Hz)</td>
</tr>
<tr>
<td>18000</td>
<td>3084</td>
<td>11040</td>
</tr>
<tr>
<td>8000</td>
<td>1159</td>
<td>5317</td>
</tr>
<tr>
<td>171</td>
<td>188</td>
<td>482</td>
</tr>
</tbody>
</table>

$\phi = 2.5$ for all BFs considered.

Table 8.1. The DRNL-filter parameters optimised at BF's of approximately 300 Hz, 8 kHz and 18 kHz (from O'Mard and Meddis, 1996).
Figure 8.3. Schematic of the DRNL filter model. Adapted from figure 1 of O'Mard and Meddis (1996).
Creating a filter-bank model

As explained above, the parameters of the DRNL model were optimised only at 3 BFs. However, a nonlinear auditory filter bank is required in order to simulate the AN response to broad-band stimuli, as will be done in the next chapter. An approximate auditory nonlinear filter bank may be generated by interpolating to other BFs the parameter values optimised at 300 Hz, 8 kHz and 18 kHz. The following Eqs., suggested by O'Mard and Meddis\(^1\), provide a tentative filter bank:

\[
\begin{align*}
\omega_{ij}^N(\omega_{bf}) &= \omega_{bf} \\
\beta^N(\omega_{bf}) &= 171.59 + 9.2726 \cdot 10^{-2} \omega_{bf} + 3.8375 \cdot 10^{-6} \omega_{bf}^2 \\
\omega_{ij}^W(\omega_{bf}) &= 372.35 + 0.63843 \omega_{bf} - 2.5234 \cdot 10^{-6} \omega_{bf}^2 \\
\beta^W(\omega_{bf}) &= 191.85 + 8.3926 \cdot 10^{-2} \omega_{bf} + 1.5699 \cdot 10^{-5} \omega_{bf}^2 \\
\alpha^W(\omega_{bf}) &= -18.364 - 1.1438 \cdot 10^{-3} \omega_{bf} + 1.7108 \cdot 10^{-8} \omega_{bf}^2
\end{align*}
\]

These equations were derived using a polynomial fitting procedure. The BF for the compound DRNL filter, \(\omega_{bf}\), was defined as the centre frequency for the narrow filter (Eq. 8.4).

It should be pointed out that the interpolation has been done only for those parameters which follow a constantly increasing or decreasing trend with BF (see Table 8.1). As for the remaining parameters (\(v\) and \(\Theta_{th}\)), special consideration must be given. With regard to the compression exponent, \(v\), it will be assumed to have the same value for all BFs in the filter bank. This approximation is introduced after the slight variability in the optimised values of \(v\) at 300 Hz, 8 kHz and 18 kHz (see Table 8.1).

As for the compression threshold, \(\Theta_{th}\), its value varies considerably with BF. Indeed, its value at 300 Hz is much higher (82 dB SPL) than its value at 8 and 18 kHz (20 and 29.5 dB SPL respectively). This large difference may be attributed in part to the pre-emphasis filter, which attenuates the signal by as much as 30 dB at 300 Hz with respect to its contribution at frequencies above 3 kHz (see Fig. 8.2). This means, for instance, that a 300 Hz signal has to be 30 dB SPL more intense than an 8 kHz signal for them to be equally compressed. As a result, the compression threshold must

---

1 The filter bank equations were suggested by O'Mard and Meddis in a personal communication.
be at least 30 dB higher at 300 Hz than at frequencies above 3 kHz. However, this only accounts for 30 of the 62 dB difference between the compression threshold at 300 Hz and 8 kHz. The remaining 32 dB difference may be attributed either to an error in the identification of $\Theta_{Th}$ in the experimental data or to an untypical set of experimental data. After all, at 300 Hz BF, the DRNL parameters were optimised (O'Mard and Meddis, 1996) by fitting the model to experimental filter shapes at high stimulus intensities (from Cooper and Rhode, 1995), where the compression threshold cannot be easily identified. Ideally, the DRNL filter should have been fitted to experimental BM I/O functions at 300 Hz BF where the compression threshold should be more easily recognised. At 8 and 18 kHz, however, the DRNL parameters were optimised to fit BM I/O functions for a wide range of stimulus intensities and, therefore, the calculated value of the compression threshold is more reliable.

The fact that the optimised $\Theta_{Th}$ is very different at low BFs does not represent a serious problem to us considering that the ultimate purpose of the model presented here is to simulate the AN rate response of pinna-based spectral cues, which occur at frequencies above 4-5 kHz (see Chapters 3 to 6). For this reason, we decided to keep the value of the compression threshold, $\Theta_{Th}$, constant for all BFs in the filter bank, and close to its optimised values at 8 and 18 kHz, for which the DRNL model was fitted more reliably to experimental BM I/O functions.

The values of $v$ and $\Theta_{Th}$ were allowed to vary slightly for fitting specific sets of experimental data, as there is experimental evidence suggesting that these parameters may vary across animals (see below). The actual values of $v$ and $\Theta_{Th}$ used in each case will be given below (see Model evaluation section below).

Since the shape of AN rate-intensity curves is largely determined by BM nonlinearities (see section 8.1 Introduction), the adequacy of this filter bank is investigated below by comparing simulated and experimental AN rate-intensity curves for stimulus at and off BF, for BFs other than 300 Hz, 8 kHz or 18 kHz.

**Nonlinear phenomena simulated by the DRNL model**

The DRNL model simulates a diversity of nonlinear-related phenomena at the BM level (O'Mard and Meddis, 1996). From our point of view, the most important are BM I/O
functions, because they determine the shape of the AN RI curves for stimuli at- and off-BF; and distortion products and two-tone suppression, because they are likely to distort the original stimulus spectrum (see Introduction and Discussion sections).

1. **Input/Output functions.** The DRNL filter model simulates experimental BM I/O functions (see O'Mard and Meddis, 1996). An example of the kind of I/O functions generated by the model is shown below in Fig. 8.16a. Notice that the model simulates the fact that for stimuli at and above BF the I/O function is linear only up to a threshold intensity (compression threshold) above which the slope of the I/O function is less than unity. For stimuli below BF, the I/O of the curve becomes more linear the lower the frequency. This behaviour coincides with the experimental observations (see Introduction above).

2. **Two-tone suppression.** Fig. 8.4 shows a comparison between two-tone suppression effects at 8 kHz BF for the DRNL model and AN data. The figure was adapted from figure 6 of O'Mard and Meddis (1996). The model was evaluated with a similar paradigm as the experimental data. A 21 dB SPL probe tone was played continuously at BF while a test tone was varied in frequency and intensity. The different shadings in the figure represent the output intensity responses evoked by the test stimulus. The dark-shading represents the suppression areas; that is, the conditions in which the presence of the test tone is able to reduce the response of the system to the probe tone with respect to the condition in which the probe tone is presented alone. The light-shading represents the excitation area; that is, the conditions in which the presence of the test tone is able to enhance the system's response to the probe tone with respect to the condition in which the probe tone is presented alone.

3. **Combination tones.** Fig. 8.5 shows a comparison between the experimental combination tone \( (2f_1-f_2) \) and the prediction by the DRNL filter bank. The figure has been adapted from O'Mard and Meddis (1996, Fig. 7). In this case, two tones \( (f_1 = 36 \text{ kHz}, \text{ and } f_2 = 38 \text{ kHz}) \) were presented simultaneously to the DRNL model at 34 kHz BF. The intensity of \( f_2 \) was varied while the intensity of \( f_1 \) was held at 81 dB SPL. The fast-Fourier transform of the output signal from the DRNL model was then calculated and the levels of the output primary tones, \( f_1 \) and \( f_2 \), and the combination tone \( (2f_1-f_2) \) were plotted as a function of the intensity of the input \( f_2 \).
Figure 8.4. Two-tone suppression phenomena at 8 kHz BF (adapted from figure 6 of O'Mard and Meddis, 1996). (a) The DRNL model response. (b) Re-plotted from auditory-nerve data by Arthur et al. (1971). The dark-shading represents suppression areas, whereas the light-shading represents the excitation area.
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

Figure 8.5. Two-tone combination tones (adapted from figure 7 of O'Mard and Meddis, 1996). BF = 34 kHz, $f_1 = 36$ kHz, and $f_2 = 38$ kHz. (a) DRNL model response. (b) Experimental data for the cat basilar membrane (figure 5a of Cooper and Rhode, 1995).
Notice that the results presented in 8.5 were obtained with the filter bank for a BF other than those at which the DRNL parameters were specifically optimised. The agreement between the experimental and the model results supports the validity of the DRNL filter bank at least for high BFs (see Discussion below).

### 8.2.4 Stage (4): BM displacement to IHC stereocilia displacement coupling

The motion of the basilar membrane produces, in turn, the motion of the IHC stereocilia. Shamma et al. (1986) proposed a simple relationship which relates the displacement of IHC stereocilia with the displacement of the basilar membrane. The relationship is as follows:

$$ \tau_c \frac{\partial u}{\partial t} + u = \tau_c C \frac{\partial w}{\partial t} \quad (8.9) $$

where: 
- $w$ is the BM displacement, 
- $u$ is the IHC stereocilia displacement, 
- $\tau_c$ is a time constant and, 
- $C$ is a coupling gain between $u$ and $w$.

This function acts as a high-pass filter. It accounts for the fact that the IHC is driven by BM velocity at low stimulus frequencies and BM displacement at higher stimulus frequencies (see Shamma et al., 1986). The trade-off frequency between the two conditions is given by $f_c = 1/2\pi\tau_c$.

In our model, the output from the DRNL filter bank was considered as $w$ in Eq. (8.9).

This stage incorporates two new parameters to the model: $\tau_c$ and $C$. The procedure followed in the optimisation of these parameters is directly linked to the optimisation of the parameters involved in the next model stage, which is explained below. The actual values of the parameters $\tau_c$ and $C$ used in the model evaluation are given in the Model evaluation section below.
The calculated value of $u$ will be used as the input to the next model stage which is concerned with simulating the IHC intracellular potential (also known as the receptor potential) as a function of stereocilia displacement.

### 8.2.5 Stage (5): IHC stereocilia displacement to IHC receptor potential coupling

The inner hair cell intracellular potential is determined by the flow of ions through ionic channels in the hair cell membrane. It is known that BM oscillations induced by the presentation of an acoustic stimulus generate, in turn, deflections of the IHC stereocilia. It is also known that stereocilia deflections determine the instantaneous intracellular voltage (for a complete review on the topic see Pickles, 1988). A simple possible explanation may be that ionic channels at the stereocilia membrane become open or closed depending on the direction of the stereocilia deflection. The proportion of open channels increases when the stereocilia are stretched and decreases otherwise.

This stage of the model is concerned with simulating the IHC intracellular potential as a function of both the IHC stereocilia displacement and the 'electrical' state of the fluid spaces that surround the hair cell. The IHC model proposed by Shamma et al. (1986) was chosen to accomplish this task. This model was designed by Shamma et al. (1986) so that it works together with the function presented in the previous stage which couples cilia-displacement with BM-displacement. As the IHC model has been fully described elsewhere (Shamma et al., 1986), only a brief overview is given here.

#### Model overview

The IHC intracellular voltage is calculated by solving the IHC equivalent electrical circuit (Shamma et al., 1986) presented in Fig. 8.6. By applying Kirchoff’s current law to that circuit the following expression is obtained:

$$
(C_a + C_b) \frac{dV}{dt} + G(u)(V - E_t) + G_K(V - E_K) = 0
$$

(8.10)

where: $E_t$ is the endocochlear potential,
\[ E_K = E_K^* + E_I R_p / (R_I + R_p) \]

\( E_K \) is the reversal potential for ionic current across the basal membrane of the IHC, and \( R_I \) and \( R_p \) are the epithelium resistances;

\[ G(u) = G_m(u) + G_a, \]

and

\( G_a \) and \( G_b \) represent the membrane capacitances associated to the apical and the basal portion of the IHC respectively.

The total conductance of the IHC membrane has two contributions: a conductance \( G(u) \) in the apical portion of the hair cell, and a potassium-dominated conductance, \( G_K \), in the basal membrane. It is assumed that all of the conductances are associated with voltage-independent ionic channels. Whence, the conductances are also voltage-independent. The apical conductance, \( G(u) \), is the sum of the mechanical conductance, \( G_m(u) \), and a leakage conductance, \( G_a \). The mechanical conductance, \( G_m(u) \), changes with stereocilia deflection, \( u \), as explained below. This produces, in turn, a change in the intracellular potential, \( V \).

As explained by Shamma et al. in their original paper, the extracellular potential at the basal portion of the hair cell is different from zero (its actual value is \( V = E_I [R_p / (R_p + R_I)] \)) but remains constant (see Shamma et al., 1986). Therefore, perturbations of the intracellular potential and of the trans-membrane potential at the basal part of the IHC become equivalent. This is an important observation that will be crucial when describing the transmitter release phenomena in the next stage of the model (see below).

It is also assumed that cilia deflection activates the ionic channels in the apical portion of the IHC directly rather than via chemical mediators. Moreover, it is assumed that there are only two possible states for the ionic channels: open and closed —i.e., activated and inactivated— (Pickles, 1988). An important consideration must be made regarding the time constant associated with channel activation-inactivation processes. There is experimental evidence that channel activation-inactivation processes follow stereocilia deflections with a delay of approximately 40 \( \mu \)s at 22°C (Corey and Hudspeth, 1979), rather than instantaneously. However, this latency time is so short that for low stimulus frequencies the activation-inactivation processes can be

---

\( ^2 \) It is assumed that \( 1/R_I \gg G, G_K \), which leads to this approximate value for \( E_K^* \) (Shamma et al., 1986).
approximated to follow stereocilia deflection instantaneously (Shamma et al., 1986). In our case, the maximum stimulus frequency considered is 18 kHz. The period associated to this frequency (55.5 μs) is still longer than the measured latency. For this reason, it has been assumed that a Boltzmann distribution can be used to represent the distribution of ion channels in the two possible states (i.e., open or closed) as an instantaneous function of the stereocilia displacement within our frequency range (1 to 18 kHz).

The mechanical conductance, $G_m(u)$, is expressed as the maximum mechanical conductance possible, $G_{\text{max}}$, (i.e., with all channels open), times the fraction of channels in the open state, $P_c(u)$. That is: $G_m(u) = G_{\text{max}} P_c(u)$. $P_c(u)$ is determined according to the Boltzmann distribution for a channel model with only two possible states. Therefore:

$$G_m(u) = \frac{G_{\text{max}}}{1 + \exp\left(\frac{\Delta G_{\text{act}}}{RT}\right)}$$

(8.11)

where: $\Delta G_{\text{act}} = G_1 - Z_\mu$ is the channel activation energy, which is modelled simply as a linear function of the stereocilia displacement,

$T$ is the absolute temperature (in Kelvin).

Note that $G_m(u)$ is partially activated at zero displacement but it is completely inactivated for large negative displacements.
Figure 8.6. Schematic diagram of a radial section of the cochlea, including the equivalent circuit of the inner hair cell (shaded box) and surrounding structures. Adapted from figure 4 of Shamma et al. (1986). $E_I =$ endocochlear potential; $R_b, R_p =$ epithelium resistances; $E_K =$ potassium reversal potential; $G_K =$ ionic channel conductance of the hair cell basal membrane; $G_m(u) =$ mechanically sensitive conductance of the hair cell apical membrane; $G_a =$ leakage conductance of the hair cell apical membrane; $u =$ cilia displacement; $V =$ IHC intracellular potential.
Transfer characteristics of the model

In order to summarise the transfer characteristics of this IHC model let us assume a sinusoidal excitation. At low stimulus amplitudes, the intracellular potential will follow the excitation waveform linearly. At high stimulus intensities, however, the IHC intracellular potential becomes saturated. The troughs of the sinusoidal signal become saturated first [because of the asymmetry of $G_m(u)$], followed by the crests at higher intensities (Shamma et al., 1986).

Regarding the temporal response of the model, the solution to Eq. (8.10) for a harmonic stereocilia deflection has the characteristics of a low-pass filter (Shamma et al., 1986). For stimulus frequencies below 4 to 5 kHz, the intracellular potential follows the stereocilia oscillation waveform. However, for higher stimulus frequencies the membrane capacitances $C_a$ and $C_b$ offer a low impedance to AC transmembrane currents. This, in effect, short-circuits the AC currents which reduces the AC voltage response in the cell (Sellick and Russell, 1980; Russell and Sellick, 1983). As a result, the receptor potential at high stimulus frequencies has only a DC component. The qualitative behaviour of this model agrees, therefore, with the experimental observations of the IHC intracellular potential (see Shamma et al., 1986; see also the Model evaluation section below).

Model parameters

The IHC receptor potential model (Shamma et al., 1986) described here has the following parameters: $E_t$, $E_K$, $q = R_p/(R_t+R_p)$, $C = (C_a+C_b)$, $G_K$, $G_0$, $G_{\text{max}}$, $\beta = \exp(-G_1/RT)$, $\gamma = Z_1/RT$. [Note that $G_0 = G(u=0)$, and determines the leakage conductance $G_a$ (Fig. 8.6)]. There are two additional parameters, $\tau_c$ and $\tau$, of the previous model stage (4). Shamma et al., (1986) proposed a full set of parameters for their model. However, they claimed that with those parameters the IHC model is only valid for stimulus frequencies up to 1.5 kHz. Therefore, a new set of model parameters had to be found with which the model reproduces experimental receptor potential data for BFs and stimulus frequencies between 4 and 18 kHz, which is the frequency range of interest to us. The criteria used in the parameter-optimisation
process, the actual parameters used in the model evaluation and the model results obtained with those parameters are given below in section 8.3: Model evaluation.

8.2.6 Stage (6): Transmitter release

The trans-membrane potential at the basal portion of IHC causes transmitter units to be released from the IHC to the IHC/AN synaptic cleft. There are no reports of any direct measurements of the relationship between the IHC membrane potential and the rate at which the transmitter is released into the IHC/AN synaptic cleft. However, studies elsewhere in the nervous system suggest that transmitter is released at a rate that is exponentially related to the deviation of the trans-membrane potential from its value at resting (Stevens, 1968). Therefore, the following function was chosen as a possible description for this stage of the transduction process:

\[ k(t) = ze^{h\theta(t)} \] (8.12)

where: \( k(t) \) is the rate of transmitter release (see below) and, \( z \) and \( h \) are parameters of the model.

In Eq. (8.12), \( \theta(t) \) represents the deviation of the membrane potential from its value at resting; that is:

\[ \theta(t) = V_m(t) - V_m^0 \] (8.13)

where: \( V_m(t) \) is the instantaneous trans-membrane potential, and \( V_m^0 \) is the resting trans-membrane potential.

We now want to express Eq. (8.12) in terms of the intracellular voltage, \( V(t) \), as this is the output from the previous model stage. Let \( V(t) \) be the instantaneous intracellular potential, and \( V_{out}(t) \) the instantaneous extracellular potential in the synaptic cleft. The trans-membrane potential can be then expressed as: \( V_m(t) = V(t) - V_{out}(t) \). At resting (i.e., in the absence of any stimulus), the trans-membrane potential is given by: \( V_m^0 = V_0 - V_{out}^0 \), where \( V_0 \) and \( V_{out}^0 \) are the intracellular and extracellular resting potentials respectively. It was pointed out in the previous stage that the extracellular potential is constant at the basal portion of the IHC. This means that \( V_{out}(t) = V_{out}^0 \).
Consequently, Eq. (8.13) can be expressed in terms of the instantaneous intracellular voltage as:

$$\vartheta(t) = V(t) - V_0$$  \hspace{1cm} (8.14)

In summary, the trans-membrane potential is equivalent to the deviation of the absolute intracellular potential, $V(t)$, from its value at resting, $V_0$. Therefore, the rate of transmitter release [Eq. (8.12)] can be expressed as:

$$k(t) = z e^{h[V(t) - V_0]}$$  \hspace{1cm} (8.15)

That is, in terms of two variables, $V(t)$ and $V_0^3$, that are directly calculated in the previous stage of the model.

The use of an exponential function for calculating the transmitter release may seem counter-intuitive, threatening extremely large transmitter values at high intensity signals. However, it must be remembered that the input to Eq. (8.15), $V(t) - V_0$, is itself saturated (see above).

It has been impossible to find a firm empirical basis for the parameters $z$ and $h$. However, an speculative explanation as to their physiological meaning, is given below in the Discussion section. Explorations of the consequences of choosing alternative values for $z$ and $h$ are also given in section 8.3, Model evaluation.

8.2.7 Stage (7): Synaptic effects

The description of the flow of transmitter across the IHC membrane into the synaptic cleft is done using the model proposed by Meddis (1986, 1988) and Meddis et al. (1990). This model has been fully explained in Chapter 7 (section 7.2.3) and, therefore, will not be explained again here.

---

3 $V_0$ must be understood as the resting potential in normal conditions, i.e., for a steady value of the parameters that determine the endocochlear potential (see Discussion section). In other words, it must be understood as a kind of reference potential.
In this case, however, the model has been implemented in a different way so that it includes the probabilistic release of individual transmitter quanta from the free store, which is believed to be a major contributor to the probabilistic nature of AN firing. In the new implementation, therefore, the quantity \( q(t) \) must be an integer and the fate of each transmitter quantum is to be decided individually and probabilistically. For example, the instantaneous probability of release of a single quantum from the free store into the cleft is given by \( k(t)dt \), where \( k(t) \) is the transmitter release rate given by Eq. (8.15). The model decides in each sampling interval whether the quantum is to be released by choosing a random number between 0 and 1. If the random number is less than \( k(t)dt \), then the quantum is released; otherwise, the quantum stays where it is. The same process is repeated for each of the \( q(t) \) available quanta.

8.2.8 Stage (8): Spike generation and refractory effects

The model of the synaptic effects described in the previous stage determine whether or not one or more transmitter quanta are present in the IHC/AN synaptic cleft at each time instant (i.e., at each sampling period, \( dt \)). It is assumed that the presence of one transmitter quantum is enough to generate a spike in the post-synaptic fibre. However, whether a spike actually occurs or not also depends on the refractory period of the fibre. A refractory period of 1 ms has been considered in our model. Therefore, a spike is produced if the time elapsed since the previous spike is longer than 1 ms and there is at least one transmitter quantum in the synaptic cleft.

8.3 MODEL EVALUATION AND PARAMETERS SELECTION

This section is concerned with presenting the response of the model to pure tone stimuli. Considering that the ultimate purpose of this model is to examine the AN rate representation of location-dependent spectral features immersed in HRTFs, our efforts have focused on optimising the model response at BFIs above 4 kHz.

An appropriate set of parameters is proposed for each stage of the model. In some cases, the consequences of using alternative parameters are examined. The model response is compared with available experimental data at various stages of the
transduction process. Firstly, the response of the receptor potential model is investigated. A set of parameters is provided with which the response of the receptor model is comparable to experimental data at higher BFs and for a range of stimulus frequencies much wider than the validity range (f<1.5 kHz) claimed by its authors (Shamma et al., 1986). Secondly, the simulated AN response is investigated. A set of parameters is presented which the model uses to simulate typical RI curves of LSR, MSR and HSR AN fibres for stimulus frequencies at and off-BF.

8.3.1 IHC intracellular receptor potential

It has been explained above that the low-pass filtering effect of the IHC apical membrane causes severe attenuation of the AC component of the receptor potential for stimulus frequencies above 4 kHz (Sellick and Russell, 1980; Russell and Sellick, 1983). Therefore, the response of the IHC receptor potential (and hence of the AN) to HRTF location-dependent spectral features must be determined almost exclusively by the DC component of the receptor potential, as location-dependent spectral features only occur at frequencies above 4 kHz. For this reason, our starting point was to search for an adequate set of parameters to simulate the dependence of the DC component of the receptor potential on stimulus intensity. Then, the model response to lower stimulus frequencies was also investigated by comparing experimental and simulated AC/DC ratios as a function of stimulus frequency. In both cases, the BF of the cell was 18 kHz.

DC receptor potential as a function of stimulus intensity

Fig. 8.7a shows a comparison between the simulated and the experimental DC component of the receptor potential as a function of stimulus intensity (DC-I curve). The experimental data has been taken from figure 6a of Patuzzi and Sellick (1983). The evaluation procedure of the model was equivalent to the experimental paradigm employed by Patuzzi and Sellick. The cell BF was 18 kHz. 25-ms tone bursts were used as the stimulus (dt = 0.01 ms). Two stimulus frequencies were tested: 18 kHz and 7 kHz. The stimulus intensity was varied from 0 to 110 dB in 10-dB steps. In the model, the DC component was calculated as the difference between the half-peak AC component of the receptor potential waveform and the resting potential (see Fig. 8.9).
Model parameters. The model parameters were optimised by the automatic method of simulated annealing (Kirkpatrick et al., 1983) to fit the experimental DC-I curve at BF. The calculated set of parameters are given in Table 8.2. Notice that the value of some parameters (those marked with an asterisk) differs from the value proposed by the authors of the corresponding model. With regard to the DRNL filter model, it was found that $\Theta_{Th} = 15.47$ dB SPL is more adequate in our case, than the 29.5 dB SPL proposed by O'Mard and Meddis (1996) —see Figs. 8.7c and 8.7d—. As for the receptor potential parameters, it was found to be impossible to fit experimental DC-I curves at 18 kHz BF with the parameters proposed by Shamma et al., (1986). Indeed, the range of validity of the receptor potential model claimed by Shamma et al. only extends up to 1.5 kHz. In our case, however, the model is to be used for a much wider range of stimulus frequencies (4 kHz < $f$ <18 kHz). For this reason, most parameters were allowed to vary with no particular restrictions in the optimisation process.

With the new set of parameters, the model is able to simulate a diversity of receptor potential data for a wide range of stimulus frequencies. For instance, it simulates experimental DC-I curves at 18 kHz BF for stimulus frequencies of 7 and 18 kHz (as shown in Fig. 8.7a) and AC/DC ratios as a function of stimulus frequency (see below). Additionally, the new parameters generate a resting potential ($V_0 = -36.7$ mV) within the experimental range (-25 to -45 mV) measured by Russell and Sellick (1978). The same set of receptor potential parameters will be employed, unless otherwise stated, in the model evaluations presented in the next sections.
Figure 8.7. (a) Experimental and simulated DC component of the receptor potential as a function of stimulus intensity for two stimulus frequencies (18 and 7 kHz). The BF of the DRNL filter is 18 kHz. The symbols represent two sets of experimental data (adapted from figure 6a of Patuzzi and Sellick, 1983). The lines represent the model results obtained with the parameters given in Table 8.2 (i.e., $\alpha_{PE} = -29$ dB SPL and $\Theta_{Th} = 15.47$ dB SPL). The stimulus frequency is indicated by the numbers by each line. (b) DRNL filter output for a pure-tone stimulus at BF (continuous line), and off-BF (long-dashed line), with the set of parameters given in Table 8.2. The BF of the filter was 18 kHz. The numbers by each line indicate the stimulus frequency. The thin dotted-line illustrates a linear relationship. (c) and (d) Same as (a) and (b) but with $\alpha_{PE} = 0$ dB SPL and $\Theta_{Th} = 29.5$ dB SPL; i.e., with the DRNL parameters originally proposed by O’Mard and Meddis (1996).
Experimental and modelled DC-I curves. Although the model parameters were only optimised to fit the DC-I curve for a stimulus at BF, they also generate a good model response for a stimulus (7 kHz) below BF. This supports the adequacy of the DRNL filter shape at least for a BF of 18 kHz.

Patuzzi and Sellick (1983) established a correspondence between the shape of DC-I curves and the BM I/O functions. This correspondence is equally found in the model. Indeed, the shape of the model DC-I curve (Fig. 8.7a) is directly related to the I/O functions of the DRNL filter bank, which is shown in Fig. 8.7b. Notice that the I/O curve is nonlinear for a stimulus at BF, but is linear for a 7 kHz stimulus. The compression 'elbow' of the nonlinear I/O function occurs at 44.47 dB SPL. This compression elbow corresponds to the break-point in the DC-I curve at the same intensity (indicated by an arrow in Fig. 8.7a). In our model, the compression elbow of the DRNL I/O curve is determined as \( \Theta_T + \alpha_{PE} \), where \( \Theta_T \) is the compression threshold of the DRNL filter and \( \alpha_{PE} \) is the pre-attenuation included in the pre-emphasis filter.

We have found that the pre-attenuation parameter, \( \alpha_{PE} \), is necessary if the model is to simulate the fact that in many cases the DC threshold occurs at stimulus intensities other than 0 dB SPL (Russell and Sellick, 1978; Patuzzi and Sellick, 1983). In other words, it accounts for the variation in cochlear sensitivity across animals. For this reason, \( \alpha_{PE} \) will be allowed to vary when fitting different sets of experimental data. For instance, the DC threshold in the experimental data shown in Fig. 8.7a occurs at around 29 dB SPL, which explains that the optimum value of \( \alpha_{PE} \) is -29 dB SPL in this case. If the pre-attenuation parameter is not considered, the DC threshold will be exclusively determined by the DRNL filter output\(^4\), which at 18 kHz BF generates DC thresholds at around 0 dB SPL as shown in Fig. 8.7c. The physiological meaning of the pre-attenuation parameter is discussed below in the Discussion section.

\(^4\) Notice that the pre-attenuation parameter, \( \alpha_{PE} \), was not included in the original implementation of the DRNL model by O'Mard and Meddis (1996).
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

Outer-/Middle-ear filter (Pre-emphasis filter)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$ filter order</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha_{\text{PE}}$ pre-attenuation (dB SPL)</td>
<td>-29</td>
</tr>
<tr>
<td>$f_{\text{lc}}$ lower cut-off frequency (Hz)</td>
<td>1000</td>
</tr>
<tr>
<td>$f_{\text{uc}}$ upper cut-off frequency (Hz)</td>
<td>30000</td>
</tr>
</tbody>
</table>

DRNL filter bank

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta_{T_h}$ compression threshold (dB SPL)*</td>
<td>15.47</td>
</tr>
<tr>
<td>$\gamma$ compression exponent</td>
<td>0.63</td>
</tr>
</tbody>
</table>

IHC Receptor potential

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_t$ endocochlear potential (V)</td>
<td>0.1</td>
</tr>
<tr>
<td>$E_K$ reversal potential (V)*</td>
<td>-0.049</td>
</tr>
<tr>
<td>$R_p/(R_1+R_p)$ reversal potential correction</td>
<td>0.04</td>
</tr>
<tr>
<td>$C=(C_a+C_b)$ total capacitance (F)*</td>
<td>6.21e-12</td>
</tr>
<tr>
<td>$G_0$ resting conductance (S)*</td>
<td>8.68e-10</td>
</tr>
<tr>
<td>$G_K$ potassium conductance (S)*</td>
<td>1.43e-8</td>
</tr>
<tr>
<td>$G_{\text{max}}$ max. mechanical conductance (S)*</td>
<td>3.98e-9</td>
</tr>
<tr>
<td>$\beta=\exp(-G_1/RT)$ beta constant</td>
<td>0.25</td>
</tr>
<tr>
<td>$\tau=Z_1/RT$ gamma constant (m$^{-1}$)</td>
<td>1.0e+7</td>
</tr>
<tr>
<td>$\tau_c$ BM/cilia time constant (s)*</td>
<td>6.56e-4</td>
</tr>
<tr>
<td>$C$ BM/cilia coupling constant (dB)*</td>
<td>-177.5</td>
</tr>
</tbody>
</table>

Table 8.2. Model parameters. The asterisks denote those parameters whose value is different from the value proposed by the authors of the corresponding model.

AC/DC ratio as a function of stimulus frequency

Another way of characterising the intracellular responses of IHCs is by representing the ratio between the AC and the DC components of the receptor potential as a function of stimulus frequency (Palmer and Russell, 1986). Fig. 8.8 shows a comparison between experimental and simulated AC/DC ratios. Two sets of experimental data (re-plotted from Palmer and Russell, 1986, Fig. 10) are represented to illustrate the variability in experimental measurements. The evaluation procedure of the model was equivalent to the experimental paradigm employed by Palmer and Russell. In the model, the BF of the cell was set to 18 kHz. 50-ms tone bursts with a 5-ms rise
period were used as stimulus ($dt = 0.01$). No fall period was applied. The frequency of the tones was varied from 100 Hz to 4.2 kHz logarithmically. The intensity of the tones was 120 dB SPL in all cases. The model response was obtained with the same parameters (given Table 8.2) employed to evaluate the DC-I curve above, with the only exception of the pre-attenuation parameter, $\alpha_{PE}$, which was set to 0 dB SPL in this case.

In the model, the AC component was calculated as the difference between the voltage at the peaks and the troughs of the receptor potential waveform; and the DC component was calculated as the difference between the half-peak AC component and the resting potential. This is illustrated in Fig. 8.9 which also shows the simulated receptor potential waveforms in response to a 120 dB SPL pure tone of various frequencies. Notice how the AC component of the receptor potential becomes more attenuated with increasing stimulus frequency, whereas the DC component continues rising (see above; also see Palmer and Russell, 1986).

---

5 Notice that the experimental data of Palmer and Russell (1986) were obtained with an 80 dB SPL stimulus, instead of the 120 dB SPL required by the model. However, one of the authors (A. Palmer) acknowledged in a personal communication with us that in many cases higher stimulus intensities (around 100 to 130 dB SPL) are required to generate AC/DC ratios like those presented in Palmer and Russell (1986).

6 Notice that only 40 ms of the full 50 ms of the signal duration are shown.
Figure 8.8. A comparison between experimental and simulated AC/DC ratios as a function of stimulus frequency. The experimental curves (black symbols) have been adapted from figure 10 of Palmer and Russell (1986). The model response is given by the white squares linked by a thick line. The BF of the cell was 18 kHz. The model parameters are given in Table 8.2 except $\alpha_{PE}$, which was set to 0 dB SPL (see text for details).
Figure 8.9. Simulated intracellular receptor potential waveforms in response to 120 dB SPL tones of various frequencies. The tone frequency is indicated (in Hz) by the side of each curve. The amplitude scale is 35 mV between ticks. The model was evaluated with the parameters given in Table 8.2. The resting potential was \(-36.7\) mV in this case.
8.3.2 AN response: LSR, MSR and HSR rate-intensity curves for stimuli at BF

As explained above, this model was developed with a view to using it for simulating the AN rate response to location-dependent spectral features present in HRTF stimuli. Therefore, because HRTF stimuli are broad-band, the model has to be able to simulate RI curves for stimuli both at and off BF, for a range of BFs between 4 and 17 kHz. This section is concerned with presenting a set of parameters with which the model simulates experimental RI curves of the three types of AN fibres (LSR, MSR and HSR) for stimuli at BF. The next section examines the model response in the case of stimuli off-BF.

Model evaluation procedure and results

Fig. 8.10 shows a comparison between experimental (Fig. 8.10a) and simulated (Fig. 8.10b) RI curves of LSR, MSR and HSR fibre types. The experimental data correspond to three representative RI curves of the several presented in figure 2 of Winter et al., (1990). The exact BF of those three units is not given in their report, but it is specified that they oscillate between 16 and 24 kHz. Therefore, an intermediate BF of 18 kHz was assumed in the model evaluation. The simulated RI curves were calculated by stimulating the model with 100 ms tone bursts. The stimulus frequency was 18 kHz. The stimulus intensity was varied from 0 to 100 dB SPL in 5-dB steps. For each intensity, AN spikes were collected over a total of 50 tone bursts. Unless otherwise stated, all simulated discharge rates presented in this chapter (those of Fig. 8.10b in particular) correspond to the average number of spikes over the last half of the stimulus burst duration (i.e., adapted rate). The parameters used to evaluate the model are described below.

It can be seen in Fig. 8.10 that the modelled RI curves show the most significant characteristics of the three types of fibres. For example, the model simulates that LSR fibres have the highest threshold of the three fibre types, which is usually 10 to 20 dB SPL above the threshold of MSR fibres (Liberman and Kiang, 1978; Liberman, 1988; Winter et al., 1990; Yates et al., 1990). Additionally, the model simulates the characteristic sloping saturation of MSR fibres (Sachs and Abbas, 1974; Winter et al.,
1990). This is only possible thanks to the incorporation of the nonlinear model of the BM. Indeed, the breakpoint in the slope of the RI curve (indicated by arrow in Fig. 8.10b) corresponds to the compression 'elbow' observed in the I/O function of the DRNL filter (Fig. 8.10c) which occurs within the dynamic range of the fibre (Yates et al., 1990). Notice, however, that the RI curve of the modelled MSR fibre saturates at stimulus intensities above 70 dB SPL. This does not happen in the experimental RI curve shown in Fig. 8.10a. However, MSR AN fibres showing this type of saturating RI curve have been indeed observed experimentally in guinea pigs (Müller and Robertson, 1991) and in cats (Liberman, 1988). Winter et al., (1990) pointed out that, in contrast to their MSR and LSR fibres measured in guinea-pigs, all MSR and LSR fibres measured by Liberman in cats (Liberman, 1988) show saturation or sloping-saturation, just as our modelled MSR fibre does.

Model parameters: Differences between LSR, MSR and HSR fibres

The model was evaluated with the same parameters used to fit the receptor potential data (given in Table 8.2). In this case, however, the pre-attenuation was set to −15 dB SPL. Fig. 8.10d shows the DC and AC components of the receptor potential obtained with these parameters. The parameters for the synaptic effects and transmitter release models are given in Table 8.3. The parameters for the synaptic effects model were chosen following the guidance given in Meddis et al. (1990), and are the same for all three fibre types (LSR, MSR and HSR).

\[\alpha_{pR} = -15 \text{ dB SPL}\] because the rate threshold occurs at a stimulus intensity of approximately 15 dB SPL for the HSR fibre.
Figure 8.10. A comparison between experimental and simulated RI curves for stimuli at BF. (a) Representative RI curves of HSR, MSR and LSR fibres [adapted from figure 2 of Winter et al. (1990)]. The BF of these units varies from 16 and 24 kHz. (b) Simulated RI curves of LSR, MSR and HSR fibres. The model results were obtained for an 18 kHz BF, with the parameters shown in Tables 8.2 and 8.3 (see text for details). (c) Input-output function of the DRNL filter bank model with the same parameters. The thin dotted line illustrates linear behaviour. (d) Simulated AC and DC components of the receptor potential as a function of stimulus intensity for the same parameter set.
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

<table>
<thead>
<tr>
<th>Synaptic effects model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>Max. No. of transmitter packets in free pool ($s^{-1}$).</td>
</tr>
<tr>
<td>$y$</td>
<td>Replenishment rate ($s^{-1}$).</td>
</tr>
<tr>
<td>$l$</td>
<td>Loss rate ($s^{-1}$).</td>
</tr>
<tr>
<td>$x$</td>
<td>Reprocessing rate ($s^{-1}$).</td>
</tr>
<tr>
<td>$r$</td>
<td>Recovery rate ($s^{-1}$).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transmitter release</th>
<th>LSR</th>
<th>MSR</th>
<th>HSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$ Transmitter release constant ($V^{-1}$).</td>
<td>1000</td>
<td>900</td>
<td>800</td>
</tr>
<tr>
<td>$z$ Transmitter release constant ($s^{-1}$).</td>
<td>0.004</td>
<td>0.5</td>
<td>11.75</td>
</tr>
</tbody>
</table>

Table 8.3. Set of parameters for the synaptic effects model (Meddis, 1986, 1988) and the transmitter release function. These parameters, together with those provided in Table 8.2, provide a complete set of parameters for simulating LSR, MSR and HSR fibres at 18 kHz BF. Note that only $h$ and $z$ vary across fibre types.

An important property of our model is that all three fibre types, LSR, MSR and HSR, can be simulated by varying only the parameters $h$ and $z$ of the transmitter release function. Fig. 8.11 illustrates the effects of varying $h$ and $z$ individually, keeping all the other parameters constant. It can be seen in this figure that $z$ controls mainly the spontaneous rate of the fibre, and also determines the rate threshold to some extent. The smaller the value of $z$, the lower the spontaneous rate of the simulated fibre (see Fig. 8.12). This is not at all surprising since a closer look to the transmitter release function [Eq. (8.15)] reveals that the rate of transmitter release at resting is exclusively determined by $z$ (see Discussion below). When the cell is driven, decreasing $z$ has the effect of decreasing the overall firing rate of the fibre, which explains the increase in rate threshold. However, a very small value of $z$ will generate unrealistically small discharge rates even for relatively high stimulus intensities (compare, for instance, the discharge rates at 100 dB SPL for $z=0.004$ in Fig. 8.11a and the experimental LSR fibre of Fig. 8.10a). In this case, increasing $h$ has the effect of increasing the discharge rate at high stimulus intensities without decreasing the rate threshold substantially. Indeed, as seen in Fig. 8.11b, $h$ has the effect of increasing the slope of

---

8 As a matter of fact, the rate threshold is not exclusively determined by $z$. The parameters of the synaptic effects model (Meddis, 1986, 1988) also influence the rate threshold (Meddis et al., 1990).
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

the RI curve without producing a substantial change in the rate threshold. The larger the value of $h$, the steeper the RI curve. Notice that both $h$ and $z$ control the dynamic range of the fibre. However, $z$ is usually determined by the fibre's SR and threshold, which leaves $h$ as the main parameter to control the dynamic range of the fibre.

The above description of the individual influence of $h$ and $z$ on modelled RI curves explains the values of $h$ and $z$ proposed in Table 8.3 for simulating LSR, MSR and HSR fibres. The final values of $h$ and $z$ given in Table 8.3 were obtained after extensive testing by 'hill-climbing' method of varying one parameter at the time and noting the effect. If the effect was beneficial the new parameter was adopted as the starting point; otherwise, it was necessary to revert to the previous value and make a different change.

In summary, we have found that the type of AN fibre is exclusively determined by the parameters $h$ and $z$ of the transmitter release function. Therefore, the model suggests that the difference between LSR, MSR and HSR fibres lies simply in the rate at which transmitter is released into their corresponding clefts. We have found that $h$ has to be gradually larger for HSR, MSR and LSR fibres, whereas $z$ has to decrease gradually from LSR to HSR fibres. The physiological meaning of these results is discussed at length in the Discussion section below. Experimental evidence is provided below to substantiate our results (Sewell, 1984).

**Simulated RI curves across BFs**

A set of parameters for simulating LSR, MSR and HSR fibres at 18 kHz BF has been presented in the previous sections. This section is concerned with investigating the response of the model at lower BFs. From the point of view of simulating the AN rate response to location-dependent HRTF features, the lowest BF of interest is 4 kHz.

The model was evaluated at 4 kHz BF with the same set of parameters as for 18 kHz BF —see Tables 8.2 and 8.3—. The resultant RI curves are shown in Fig. 8.13a. Notice that we made no attempt to fit the model at 4 kHz BF to a specific set of experimental RI curves, as that would require different values of $h$ and $z$ from those given in Table 8.3 (see below). Our intention was simply to check that the simulated RI curves at 4 kHz BF show the most important characteristics of LSR, MSR and
HSR fibres. In this sense, the resultant RI curves at 4 kHz BF are similar to the experimental RI curves (at 18 kHz) shown in Fig. 8.10a. However, the simulated RI curves at 4 kHz have a steeper slope than those at 18 kHz. This reduces the characteristic sloping-saturation of the simulated MSR fibre and causes both MSR and LSR fibres to clearly saturate. In the model, this is a direct consequence of two factors. Firstly, the I/O function of the DRNL filter bank is slightly 'more linear' at 4 kHz (Fig. 8.13c) than at 18 kHz (Fig. 8.10c). Secondly, the AC component contributes more at 4 kHz (Fig. 8.13d) than at 18 kHz (Fig. 8.10d) and, therefore, the peak value of the receptor potential is larger at 4 kHz.

There are several ways of improving the modelled RI curves at 4 kHz BF. In practical terms, the simplest way is by decreasing either \( z \), or \( h \) (or both) at 4 kHz with respect to their values at 18 kHz BF. As a matter of fact, there is experimental evidence (see Discussion section) indicating that \( h \) must be smaller for lower BFs. For this reason, we decided to reduce the value of \( h \) for simulating the RI curves at 4 kHz. Possible values of \( h \) at 4 kHz BF are given in Table 8.4. Fig. 8.13b shows the resultant RI curves at 4 kHz with these parameters. A comparison between Figs. 8.13a and 8.13b shows that reducing \( h \) at 4 kHz produces the effect of reducing the slope of the simulated RI curves and, therefore, generates more realistic RI curves, particularly in the case of MSR and LSR fibres.

<table>
<thead>
<tr>
<th>BF (kHz)</th>
<th>LSR ( h ) (V(^{-1}))</th>
<th>LSR ( z ) (s(^{-1}))</th>
<th>MSR ( h ) (V(^{-1}))</th>
<th>MSR ( z ) (s(^{-1}))</th>
<th>HSR ( h ) (V(^{-1}))</th>
<th>HSR ( z ) (s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>900</td>
<td>0.004</td>
<td>800</td>
<td>0.5</td>
<td>700</td>
<td>11.75</td>
</tr>
<tr>
<td>18</td>
<td>1000</td>
<td>0.004</td>
<td>900</td>
<td>0.5</td>
<td>800</td>
<td>11.75</td>
</tr>
</tbody>
</table>

Table 8.4. Values of the parameters \( h \) and \( z \) for simulating LSR, MSR and HSR at 4 and 18 kHz BF. Notice that the values at 18 kHz coincide with those of Table 8.3. Notice that \( z \) does not change with BF.

In summary, allowing \( h \) to be smaller for lower BFs, the model simulates realistic RI curves of LSR, MSR and HSR AN fibres at 4 and 18 kHz BF. In fact, similar results (not shown here) are obtained for intermediate BFs.

The model RI curves at 4 kHz are more realistic if a lower \( h \) is considered because: (1) the rate threshold for LSR fibres is higher, (2) the LSR fibre does not reach saturation within the considered intensity range, and (3) the sloping saturation of the MSR fibre is more visible.
Figure 8.11. The effect of varying the parameters $h$ and $z$ of the transmitter release function individually. All the other parameters of the model have been kept constant. Notice how the SR decreases by decreasing $z$, and how $h$ controls the slope of the RI curve without changing the rate threshold substantially.
Figure 8.12. Spontaneous rate in the model as a function of the parameter $z$ of the exponential transmitter release function.
Figure 8.13. (a) Modelled RI curves at 4 kHz BF using the same parameters set as for 18 kHz BF (Tables 8.2 and 8.3). (b) Modelled RI curves at 4 kHz BF for a smaller h (see Table 8.4). (c) DRNL I/O function at 4 kHz BF with the parameters given in Table 8.2, for a stimulus at BF (d) AC and DC components of the receptor potential at 4 kHz BF for a stimulus at BF.
8.3.3 AN response: Rate-intensity curves for stimuli off BF

This section is concerned with comparing modelled and experimental RI curves for stimulus frequencies below and above BF. The comparison is done for a HSR and a MSR AN fibre, both with a BF of 16 kHz. The experimental data were taken from figures 2 and 3 (HSR and MSR fibres respectively) of Yates et al. (1990).

The procedure was as follows. Firstly, a set of parameters was obtained to simulate the RI curve of each fibre type at BF. Secondly, the model was evaluated for stimulus frequencies below and above BF using the parameters obtained at BF. In the model, spikes were collected for 50 tone bursts of 100 ms of duration ($dt = 0.01$ ms). The stimulus intensity was varied from 0 to 60 dB SPL in 5-dB steps. The discharge rate shown in the simulated RI curves corresponds to the average discharge rate over the last half of the tone-burst duration.

Model parameters

A useful comparison between the model performance and the experimental data for stimuli off-BF can only be made if the modelled RI curves at BF fit the experimental data to a good approximation. In our case, this meant having to obtain a new set of parameters for simulating the experimental RI curves at BF of Yates et al. (1990), instead of using the existing parameters given in Tables 8.2 and 8.3. However, to be self consistent, only four parameters of those given in Tables 8.2 and 8.3 were allowed to differ from those previously set in the model: the pre-attenuation, $\alpha_{PE}$, the compression threshold, $\Theta_{Th}$, and $h$ and $z$ of the transmitter release function. The reason we allowed these parameters to vary is because they are likely to be different for different animals (in the case of $\alpha_{PE}$ and $\Theta_{Th}$) and for specific fibres (in the case of $h$ and $z$). The new values of these parameters are given in Table 8.5. Notice that only $h$ and $z$ are different between the MSR and HSR fibres.
Table 8.5. Parameters for fitting the RI curves of MSR and HSR fibres shown in figures 2 and 3 of Yates et al., (1990). The other model parameters are the same as those given in Tables 8.2 and 8.3. Notice that only $h$ and $z$ vary for the two fibre types.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSR</th>
<th>HSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{PE}$ (dB SPL)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\Theta_{Th}$ (dB SPL)</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>$h$ (V$^{-1}$)</td>
<td>750.0</td>
<td>650.0</td>
</tr>
<tr>
<td>$z$ (s$^{-1}$)</td>
<td>0.27</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Results: off-BF RI curves

A comparison between the model results and the experimental RI curves is shown in Figs. 8.14 (HSR fibre) and 8.15 (MSR fibre). The I/O functions of the DRNL model and the simulated DC receptor potential as a function of stimulus intensity corresponding to the model RI curves are shown in Fig. 8.16.

For a stimulus at BF, the modelled RI curves are in good agreement with the experimental data. For stimuli below and above BF, the model response is qualitatively and quantitatively correct for both fibres (see Yates et al., 1990) for stimulus frequencies close to BF. However, for the HSR fibre, the discharge rate of the model for stimulus frequencies far from the BF (12 and 20 kHz) is larger than the experimental values. Because RI curves for stimuli off BF are largely determined by the filter shape (Yates et al., 1990), a possible reason for this discrepancy is that the DRNL filter is too wide at 16 kHz BF. However, the same discrepancy is not observed for the MSR fibre, even though the DRNL filter output is identical for the two fibres (see Figs. 8.16a). Notice that the receptor potential (Fig. 8.16b) is also the same for both fibres because they have been simulated with the same receptor potential parameters. A possible explanation for these results may be that the data for the MSR and the HSR fibres were taken from different animals; in which case different DRNL parameters would be required for each fibre. This possibility, however, can not be confirmed or rejected with the information given in the original paper from where the experimental data was taken (Yates et al., 1990).
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

Figure 8.14. Experimental and modelled RI curves for a HSR AN fibre for stimuli at and off BF. The BF of the fibre is 16 kHz. The experimental data (symbols) have been adapted from figure 2 of Yates et al. (1990). The modelled RI curves are represented by the lines (see text for details). The stimulus frequency for each curve is indicated in the legend. (a) For stimulus frequencies at and below BF. (b) For stimulus frequencies at and above BF.
Figure 8.15. Experimental and modelled RI curves for a MSR AN fibre for stimuli at and off BF. The BF of the fibre is 16 kHz. The experimental data (symbols) have been adapted from figure 3 of Yates et al. (1990). The modelled RI curves are represented by the lines (see text for details). The stimulus frequency for each curve is indicated in the legend.
Figure 8.16. (a) I/O functions of the DRNL filter at 16 kHz BF. Different symbols indicate different stimulus frequencies. (b) The modelled DC component of the receptor potential as a function of stimulus intensity. These results correspond to the model RI curves shown in Figs. 8.14 and 8.15.
8.4 DISCUSSION AND CONCLUSIONS

We have presented a computer model of the peripheral auditory system which incorporates a nonlinear auditory filter bank to simulate the compression effect by the basilar membrane. The model simulates experimental IHC intracellular potential data. Additionally, (and most importantly from our point of view) the model simulates experimental RI curves of LSR, MSR and HSR AN fibres for stimuli at BF, for a range of BFs between 4 and 18 kHz. Furthermore, the model simulates approximately the right AN rate response for stimuli off BF. The proposed model also embraces two-tone suppression and distortion products which are important phenomena to take into account when modelling the AN rate response to HRTF stimuli.

8.4.1 Model parameters

The DRNL filter bank. In the model, the compression exponent of the DRNL filter bank (O'Mard and Meddis, 1996) has been assumed to be constant across the range of BFs considered (4 to 18 kHz). This is supported by the results published by Cooper and Yates (1994) who found that the slope of BM I/O functions derived from AN fibres varies very little across BFs for BFs above 4 kHz.

The receptor potential model. A crucial stage in our model is the receptor potential model developed by Shamma et al. (1986). Despite its simplicity (Shamma et al., 1986), this receptor potential model simulates a variety of experimental data for a wide range of stimulus frequencies. However, it has been required to change the value of some of the model parameters with respect to those proposed originally by Shamma et al. Notice that Shamma et al. claimed that their model was valid only up to 1.5 kHz and we have extended the valid range to higher frequencies (18 kHz).

The pre-attenuation parameter. It has been found that a pre-attenuation parameter is required in order to simulate the experimental threshold for specific sets of data (see above). In particular, it has been required to simulate the threshold of experimental DC-I curves (Fig. 8.7). It is difficult to give a physiological explanation to the pre-attenuation parameter. This would require knowing exactly what causes
Chapter 8: A computer model of LSR, MSR and HSR fibres using a nonlinear auditory filter bank

experimental thresholds to be at stimulus intensities other than 0 dB SPL in each data set. One possibility is that cochlear sensitivity varies across animals and, therefore, the pre-attenuation corresponds to a property of the cochlear filtering which has been unaccounted for in the original DRNL filter model of the BM (O'Mard and Meddis, 1996). A second possibility is that the IHCs from which the experimental data were collected were in fact damaged at the time of measuring. Russell and Sellick (1978) measured DC-I curves (in healthy cochleae) in which the threshold varied from -2 to 50 dB SPL (see table 1 of Russell and Sellick, 1978). In their measurements, those cells showing a DC-I threshold above 20 dB SPL also show a high compound action potential threshold (=15 dB SPL). Similarly, Patuzzi and Sellick (1983) measured DC-I curves in healthy cochleae—two of which are shown in Fig. 8.7a—which had thresholds at around 30 dB SPL. The compound action potential threshold for those units was around 31 dB SPL. Therefore, it appears to be a correlation between the high threshold in DC-I curves and the high compound action potential thresholds. High action potential thresholds usually indicate that the cell has been damaged during experimental manipulation. Consequently, it is likely that the high thresholds in DC-I curves and, therefore, the pre-attenuation parameter, are associated to damaged cells. However, the possibility that the pre-attenuation serves to account for the variation in cochlear sensitivity across animals can not be ruled out on the existing evidence.

8.4.2 Low, Medium and High spontaneous rate AN fibres

One of the most important properties of the proposed model is that it simulates the RI curves of LSR, MSR and HSR AN rate fibres. It has been found that the three fibre types can be simulated by varying only the parameters $z$ and $h$ of the transmitter release exponential function. This has several implications. Firstly, the three fibre types are driven by the same IHC intracellular potential in our model. Therefore, the model reflects the experimental fact that AN fibres of the three types are known to contact the same IHC (Liberman and Oliver, 1984). Secondly, the model suggests that LSR, MSR and HSR fibres differ only in the rate at which transmitter is released to the synaptic cleft of each fibre type. Therefore, it is implicit in our model that the physiological differences between HSR, MSR and LSR are presynaptic in the relation between the trans-membrane potential and the transmitter release. This is supported by the results of Liberman et al. (1990) who provided evidence of presynaptic morphological differences (see below) between the pillar- and the modiolar-side.
synapses, thus between high- and low/medium-SR fibres\(^\text{10}\) \((\text{Liberman and Oliver, 1984})\).

It has been found in the model that the value of \(z\) is lower for LSR fibres and increases progressively for MSR and HSR fibres. As for the value of \(h\), it is highest for LSR fibres and decreases progressively for MSR and HSR fibres (see Tables 8.3 and 8.5). These findings are supported by the those of Sewell \((1984)\) in cats. Sewell injected furosemide in cats to provoke variations in the endocochlear potential, and then measured the SR of AN fibres as a function of the instantaneous endocochlear potential. He found an exponential relationship between the SR of AN fibres and the endocochlear potential. This relationship was of the form: \(R_S = R_0 e^{kE_t}\), where \(R_S\) is the spontaneous rate, \(R_0\) and \(k\) are constants\(^\text{11}\), and \(E_t\) is the endocochlear potential. [Notice the resemblance between this relationship and the exponential transmitter release function in our model (Eq. 8.15); \(R_0\) would correspond to \(z\) and \(k\) to \(h\).]

Sewell found that \(R_0\) is higher for HSR fibres than for MSR and LSR fibres\(^\text{12}\), whereas \(k\) is lower for HSR fibres than for MSR and LSR fibres. These results agree qualitatively with our findings regarding the variation of \(h\) and \(z\) with the type of AN fibre.

The agreement between Sewell's findings and the required values of \(z\) and \(h\) in our model is a direct consequence of the linear relationship between the IHC intracellular potential and the endocochlear potential at resting\(^\text{13}\). At resting, our exponential transmitter release function (Eq. 8.15) can be expressed as a function of the endocochlear potential. It is very important to notice, however, that \(V_0\) in Eq. 8.15 must be understood as the resting potential in normal conditions, i.e., for a steady endocochlear potential in normal conditions (in our case, this value is \(E_T=100\ \text{mV}\), which yields a resting potential \(V_0 = -36.7 \text{ mV}\)). In other words, \(V_0\) in Eq. 8.15 does not change when the endocochlear potential is forced to vary. Therefore, \(V_0\) can be interpreted as a (constant) reference potential in the model. Otherwise, the model

\(^{10}\) Liberman and Oliver \((1984)\) found that all LSR and MSR AN fibres have synaptic sites on the modiolar side of IHC, whereas most HSR have synaptic contacts in the pillar side of the IHC.

\(^{11}\) \(R_0\) corresponds to the spontaneous rate when the endocochlear potential, \(E_t\), is zero.

\(^{12}\) The results of Sewell \((1984)\) are restricted to HSR and MSR. No results for LSR are actually presented in his report. However, because of the morphological similarities between MSR and LSR \((\text{Liberman et al., 1990; Liberman and Oliver, 1984})\) the results found by Sewell regarding MSR fibres are possibly generalisable to LSR fibres.

\(^{13}\) This linear relationship is derived from Eq. (8.10) by introducing the resting condition \(dV/dt = 0\).
would not be able to account for variation of SR with $E_t$ found by Sewell as the transmitter release rate would be determined exclusively by $z$.

The agreement between Sewell's results and the value of $h$ in our model is not only qualitative but also quantitative. When plotting the logarithm of the spontaneous rate as a function of the endocochlear potential, Sewell found slopes from 0.0078 to 0.066 (log SR/mV). We have replicated Sewell's results with our model and measured the spontaneous rate as a function of the endocochlear potential (notice that the endocochlear potential is a parameter of the receptor potential model by Shamma et al., 1986, and, therefore, can be varied). In this case, the model simulates the exponential relationship between the SR and the endocochlear potential. Additionally, the slopes predicted by the model with the values of $h$ and $z$ given in Tables 8.3, 8.4 and 8.5 are within the experimental range measured by Sewell. For instance, with the parameters $h$ and $z$ given in Table 8.3, we have found slopes of 0.032 and 0.027 log(SR)/mV, for MSR and HSR fibres respectively.

Another important conclusion by Sewell (1984) is that the rate at which the spontaneous rate increases with an increase in the endocochlear potential (i.e., the slope, $k$, of log(SR)-$E_t$ curves) is lower for lower BFs. Notice that this conclusion is based on the assumption that the primary effect of the furosemide, namely a significant reduction in the endocochlear potential, is similar throughout the cochlea. If that was the case, Sewell's finding would agree with the fact that a lower $h$ is necessary for lower BFs (4 kHz) in the model in order to generate similar discharge rates at 4 and 18 kHz BF.

The physiological meaning of $h$ and $z$ remains unclear. However, it is worth speculating about it on the existing knowledge about the IHC/AN synapse. It may be that the parameter $z$ is related to the size of the synaptic plaque\textsuperscript{14}, whereas $h$ is related to the length or the height-width ratio of the synaptic body. These speculations are substantiated by the morphological differences between HSR and LSR (and MSR) fibres reported by Liberman and Oliver (1984) and Liberman et al. (1990). Liberman et al. (1990) found that the height and height-width ratio of the synaptic body is larger for synapses in the modiolar side (i.e., for LSR fibres) than for those in the pillar side (i.e., for HSR fibres) of the IHC. Since $h$ is larger for LSR fibres than for HSR fibres, the model simulates the exponential relationship between the SR and the endocochlear potential.

\textsuperscript{14} The synaptic plaque is the portion of IHC membrane specialised in the IHC/AN synapse.
fibres, it seems reasonable to think that $h$ may reflect some property of the synaptic body. As for $z$, it may be related to the size of the synaptic plaque. We know of no data which shows a correlation between the size of the synaptic plaque and fibres' SRs. However, Liberman and Oliver (1984) concluded that the axons of HSR fibres are thicker on the average than those of MSR and LSR fibres, which suggests that the size of the synapse must be larger for HSR than for MSR and LSR. Since $z$ is larger for HSR than for MSR and LSR, it seems reasonable to think that $z$ is directly related to the synaptic size.

8.4.3 Model applications

The proposed model has been developed with a view to using it for simulating the AN rate response to HRTF-filtered stimuli (see next chapter). The fact that the model simulates two-tone suppression phenomena and the presence of distortion products is very important. These phenomena must modify the spectral content of the HRTF-filtered signal by destroying or enhancing the location-dependent spectral cues useful for sound localisation. The appearance of distortion products in a nonlinear system such as the cochlea is likely to smear the stimulus spectrum, by filling up the spectral notches. On the contrary, the existence of two-tone suppression phenomena is, at first sight, likely to enhance the spectral characteristics of the signal (Pickles, 1988, p.103). Therefore, it seems very probable that HRTF pinna-related spectral features (particularly spectral notches) at the AN are very different (and may be non-existent) from those present in the original HRTF. This rises the question of how HRTF spectral cues are signalled to the brain (if at all). This issue is investigated in the next chapter.
CHAPTER 9

Simulating auditory nerve responses to HRTF stimuli using a nonlinear auditory filter bank

9.1 INTRODUCTION

This chapter is concerned with investigating the auditory-nerve (AN) encoding of pinna-based spectral cues (see also Chapter 7). The study is done by simulating the rate response\(^1\) of AN fibres to broad-band noise stimuli filtered through human-like head-related transfer functions (HRTFs). The nonlinear model of the peripheral auditory system described in Chapter 8 is employed in the simulation.

Several issues are investigated: (1) The degree to which HRTF spectral features are reflected in the cochlear output profiles and AN rate profiles at various signal levels. (2) A comparative study between the quality of the onset and the steady-state AN rate representations of HRTF spectra. (3) A comparative study of the quality of the AN rate representations associated with low-, medium- and high-spontaneous rate (LSR, MSR and HSR respectively) fibres at various stimulus intensities. (4) The changes in the AN rate response associated with a moving stimulus switching between two different locations.

The first three issues were already investigated in Chapter 7. Two main conclusions were drawn in that case: (1) Even though the BM mechanical filtering smears the stimulus spectrum by filling up the notches and destroying the fine spectral structure, \(^1\) Pinna-based spectral features occur at frequencies (> 4 kHz) above the cut-off of phase locking phenomena in the AN response. Therefore, the AN encoding of pinna-related cues must be done exclusively in terms of the discharge rate of AN fibres (Rice et al., 1995; see also Chapter 7).
the broad spectral features (including notches) are still present in the BM response; (2) The same effect was observed at all stimulus intensities, which makes the dynamic range of AN fibres (either at the stimulus onset or in the steady state, and for LSR, MSR or HSR fibres) the only restriction for the spectral features of the stimulus to be encoded in the AN response at high stimulus intensities. However, the auditory filter bank employed in Chapter 7 was linear and, therefore, failed to simulate the effects (described in Chapter 8) associated with the nonlinear cochlear filtering. Therefore, the conclusions drawn in Chapter 7 are only valid for stimuli in the linear region of the cochlear filtering (i.e., low stimulus intensities). As explained in Chapter 8, for stimuli in the compression region, BM nonlinearities generate distortion products and are also responsible for two-tone suppression phenomena. These two phenomena have been measured in the response of AN fibres (Arthur et al., 1971; Goldstein and Kiang, 1968).

It was explained in Chapter 8 that two-tone suppression phenomena and the presence of distortion products in the AN response must distort the AN representation of the spectral features of the stimulus at signal levels in the compression region of the BM filtering. In terms of sound localisation, this means that the location-dependent spectral features available in HRTFs must be distorted, particularly at high stimulus intensities. Despite this, it is known that some pinna-related spectral features available in HRTFs are actually used as cues in sound localisation tasks. For instance, elevation-varying spectral notches are known to be important cues in elevation judgements (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Bloom, 1977; Watkins, 1978).

The work presented below attempts to shed some light on the solution to this apparent paradox by simulating the AN rate response to HRTF-filtered stimuli using the nonlinear auditory model presented in Chapter 8. The main result of this study is that the quality of the AN rate profiles decreases with increasing signal level, mainly as result of the broadening of the cochlear filters. This result suggests that sound localisation accuracy must decrease with increasing signal level in tasks where pinna-based spectral features are the only localisation cues available.

Additionally, the model is employed to make predictions about the AN response to stimuli moving from one HRTF to another. Poon and Brugge (1993a,b) studied the response of individual AN fibres as the sound source moves in space. They found
that spectral notches are encoded in AN discharge patterns. However, as discussed by Rice et al. (1995), "...the nature of the neural representation of a particular stimulus spectrum cannot be inferred from those data (i.e., data for a single AN fibre)...". Consequently, Rice et al. (1995) measured AN rate profiles for a large population of AN fibres in response to broad-band noise filtered through cat HRTFs. They found that the quality of simple AN rate profiles is very poor because of the large variability in the rate characteristics of the fibres (see Discussion below). They suggested that a way of eliminating the fibre variability is by using the rate-difference profile between current responses and responses in the immediate past. However, they measured AN rate profiles in response to stationary stimuli and, therefore, their estimated rate-difference profiles do not reflect transient changes in AN rate response as the stimulus switches between HRTFs.

An attempt is made below to use the nonlinear auditory model proposed in Chapter 8 in order to make predictions about transient changes as well as average changes in AN rate profiles as the stimulus switches from one HRTF to another. [It is important to point out that although the model parameters were optimised to fit physiological data for animals, studies currently in progress suggest that the nonlinear characteristics of cochlear filtering are very similar in humans]. The main result of this study is that the transient AN rate profiles as the stimulus switches from one HRTF to another can be interpreted (at least at moderate signal levels) as the first order differential of the two successive HRTFs. Some ways in which this differencing mechanism may aid in the estimation and detection of HRTF spectral features are discussed.

### 9.2 STIMULUS GENERATION

The stimuli were broad-band noise filtered to have the spectra of human-like HRTFs. The noise was generated as a sequence of Gaussian pseudo-random numbers sampled at a frequency of 44100 Hz. A different noise segment was generated for every model

---

2 Stuart Rosen has derived nonlinear input/output functions from psychophysical masking experiments in humans and observed that they are comparable to basilar-membrane data for chinchillas (S. Rosen, personal communication). In an independent study, Andrew Oxenham and Chris Plack have got to similar conclusions (A. Oxenham and C. Plack, personal communication).

3 The Gaussian pseudo-random numbers were generated as indicated in Hastings and Peacock (1975, p. 100).
run in order to prevent the own spectrum of the noise segment from biasing the results.

The model was evaluated for three different stimuli: two "stationary stimuli" (S1 and S2) and a "moving stimulus" (MS). The stationary stimuli, S1 and S2, were generated by convolving 100-ms noise segments with the head-related impulse responses (HRIRs) corresponding to the following spatial locations: S1 0° elevation, 0° azimuth; and S2 -20° elevation, 0° azimuth. The moving stimulus, MS, was generated by concatenating two 50-ms noise segments (NS1 and NS2), each of which had been previously convolved with the HRIRs corresponding to the source locations S1 and S2. The concatenation was done in the order NS1-NS2, so that the result was a single 100-ms noise segment which corresponds to a noise source moving abruptly from 0° to -20° elevation in the frontal median vertical plane (0° azimuth). There were no audible clicks in the moving stimulus noise segment that could be identified with the junctions between the two adjacent noise segments. In all cases, the convolution was done as explained in section 7.1 of Chapter 7 (Model description). The three stimuli, S1, S2 and MS were then applied using a 2.5-ms ramp.

The two HRIRs considered were taken from the database released by Gardner and Martin (1994, 1995) and correspond to their files L0e000a and L-20e000a. Both of them were measured at the left ear (pinna model DB-061) of a KEMAR (see Chapter 4). There was no particular reason for selecting those two HRIRs other than their spectra are considerably different and display a variety of prominent spectral peaks and notches, in the frequency region from 4 to 17 kHz. The spectra of these two HRIRs are shown in Fig. 9.1. Note that the most salient features (peaks and notches) are successively numbered in each HRTF. (Because of the approximate logarithmic spacing of the different frequency channels in the cochlea, all spectra and AN rate profiles shown below are represented as a function of the log_{10} of the frequency in Hz units —see Fig. 9.1c). Five different stimulus intensities were considered for each of the three stimuli: 20, 30, 40, 60 and 80 dB SPL (rms). These rms values correspond respectively to approximately −23.43, −13.43, −3.43, 16.57 and 36.57 dB spectrum level^4 (dB SPL/Hz). Unless otherwise stated, all intensity values referred to below correspond to average rms values.

---

^4 The spectrum level (dB SPL/Hz) was calculated by subtracting 10log_{10}(noise-bandwidth) from the dB SPL (rms) value. In our case, the noise bandwidth was 22050 Hz, corresponding to a 44.1 kHz sampling frequency.
Figure 9.1. Spectrum of the head-related impulse responses considered (a) S1, 0° elevation, 0° azimuth; and (b) S2, -20° elevation, 0° azimuth. Notice that the spectrum is represented as a function of the $\log_{10}$ of the frequency (in Hz), rather than the frequency itself. The correspondence between these two variables is shown in panel (c).
9.3 MODEL PARAMETERS, SIMULATED RATE-INTENSITY CURVES AND EVALUATION PROTOCOL

The computer model of the auditory system presented in Chapter 8 was employed to simulate the AN rate profiles. The full set of parameters employed in the model evaluation are given in Table 9.1. Those parameters were selected because they simulate the RI curves of MSR and HSR fibres for pure-tone stimuli at- and off-BF (see section 8.3, Model Evaluation and Parameter Selection, of Chapter 8).

As explained in Chapter 8, the parameters given in Table 9.1 were optimised for a BF of 16 kHz. It was suggested in Chapter 8 that a lower value of $h$ should be employed at lower BFs (4 kHz) for the model to generate similar discharge rates at 4 and 16 kHz. However, for simplicity, this suggestion has not been considered in the model evaluation presented here and $h$ has been kept constant across BFs. The main result of this decision is that discharge rates for equivalent signal levels are slightly higher at 4 kHz than at 16 kHz, particularly for MSR and LSR fibres (see below). The issue is discussed in the Discussion section.

Table 9.1. Model parameters (continues in the next page).

<table>
<thead>
<tr>
<th>Outer-/Middle-ear filter (Pre-emphasis filter)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>filter order</td>
</tr>
<tr>
<td>$a_{PE}$</td>
<td>pre-attenuation (dB SPL)</td>
</tr>
<tr>
<td>$f_{1c}$</td>
<td>lower cut-off frequency (Hz)</td>
</tr>
<tr>
<td>$f_{uc}$</td>
<td>upper cut-off frequency (Hz)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DRNL filter bank</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta_T h$</td>
<td>compression threshold (dB SPL)</td>
</tr>
<tr>
<td>$v$</td>
<td>compression exponent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IHC Receptor potential</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_i$</td>
<td>endocochlear potential (V)</td>
</tr>
<tr>
<td>$E_K$</td>
<td>reversal potential (V)</td>
</tr>
<tr>
<td>$R_p/(R_t+R_p)$</td>
<td>reversal potential correction</td>
</tr>
<tr>
<td>$C_a+C_b$</td>
<td>total capacitance (F)</td>
</tr>
<tr>
<td>$G_0$</td>
<td>resting conductance (S)</td>
</tr>
</tbody>
</table>
Chapter 9: Simulating AN encoding of pinna-based spectral cues

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_K$</td>
<td>potassium conductance (S)</td>
<td>1.43e-8</td>
</tr>
<tr>
<td>$G_{max}$</td>
<td>max. mechanical conductance (S)</td>
<td>3.98e-9</td>
</tr>
<tr>
<td>$\beta$</td>
<td>beta constant</td>
<td>0.25</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>gamma constant (m$^{-1}$)</td>
<td>1.0e+7</td>
</tr>
<tr>
<td>$\tau_c$</td>
<td>BM/cilia time constant (s)</td>
<td>6.56e-4</td>
</tr>
<tr>
<td>$C$</td>
<td>BM/cilia coupling constant (dB)</td>
<td>-177.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synaptic effects model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>Max. No. transmitter packets in free pool (s$^{-1}$)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Replenishment rate (s$^{-1}$)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Loss rate (s$^{-1}$)</td>
</tr>
<tr>
<td>$x$</td>
<td>Reprocessing rate (s$^{-1}$)</td>
</tr>
<tr>
<td>$r$</td>
<td>Recovery rate (s$^{-1}$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transmitter release</th>
<th>LSR</th>
<th>MSR</th>
<th>HSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>1000</td>
<td>750</td>
<td>650</td>
</tr>
<tr>
<td>$z$</td>
<td>0.003</td>
<td>0.27</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Table 9.1 (continued). Model parameters.

9.3.1 Rate-intensity curves for broad-band noise

The simulated RI curves presented in the previous chapter were for pure-tone stimuli. In this chapter, however, the model is evaluated for broad-band noise stimuli. It is known that RI curves for broad-band noise and pure-tone stimuli are slightly different for several reasons, the main of which is that two-tone suppression phenomena influence the discharge rate of the fibre in the case of noise stimuli (Schalk and Sachs, 1980). Since the AN rate profiles presented below were obtained for broad-band noise stimuli, it is appropriate to present the simulated RI curves for noise stimuli obtained with the above parameters.

The DRNL filter input-output (I/O) functions and the simulated AN RI curves for broad-band noise stimuli are shown in Fig. 9.2 (onset RI curves) and Fig. 9.3 (steady-state RI curves). In this case, the stimulus was a 100-ms Gaussian broadband noise generated as explained above. A post-stimulus time histogram (PSTH) was produced by running the model 50 times, employing a different noise burst each
time. From the PSTH, the onset response was calculated as the maximum rate over the first 10 ms of the noise burst, and the steady-state response as the average discharge rate over the last 50 ms of the noise burst.

Notice that the rate threshold of the noise RI curves shown in Fig. 9.3 is roughly 10 dB higher than for pure-tone stimuli (see Figs. 8.14 and 8.15). This occurs because for the same rms stimulus intensity, the energy flowing per frequency channel is lower in the case of broad-band noise stimuli.

9.3.2 Model evaluation protocol

For each stimulus condition, a PSTH (1 ms bin-width) was generated by collecting the number of AN spikes over 50 model runs. For each run, the HRTF-filtered stimuli were generated (as explained above in the Stimulus generation section) with a different 100-ms noise segment. The total number of spikes was then converted to discharge rates (in spikes/s). The process was carried out for each one of the three stimuli S1, S2 and MS at all five intensities considered: 20, 30, 40, 60 and 80 dB SPL (rms).

The results were presented in the form of rate profiles in which the discharge rate is represented as a function of the fibre BF. 50 logarithmically-spaced frequency channels (or BFs) were considered in the model evaluation within the interval 4 to 17 kHz. In the case of stationary stimuli, the onset rate profiles were calculated, unless otherwise stated, as the maximum discharge rate over the first 10 ms of the stimulus. Steady-state rate profiles were calculated as the average over the last half (i.e., the last 50 ms) of the stimulus duration. For the moving stimulus, the responses are represented as maximum-rate profiles calculated over a 10 ms period as the stimulus switches from the first to the second location (see below).

The output profiles from the DRNL filter bank were also calculated for each stimulus condition and intensity. They were calculated to represent the average profile across the 50 model runs. Because a different noise burst was used for each run, this prevents the results from being biased by the spectral characteristics of a particular noise segment.
Chapter 9: Simulating AN encoding of pinna-based spectral cues

Figure 9.2. Model results at 4 (black squares) and 16 kHz (white squares) BF for broad-band noise stimuli. (a) DRNL filter input-output function. The dotted lines illustrate a linear behaviour. (b)-(d) Onset RI curves: (b) HSR fibre, (c) MSR fibre and (d) LSR fibre. The intensity in the X-axis is rms (See text for details).
Figure 9.3. Simulated steady-state RI curves at 4 and 16 kHz BFs for broad-band noise stimuli. (a) HSR fibre; (b) MSR fibre and, (c) LSR fibre. The intensity in the X-axis is rms, (See text for details). Notice that the rate threshold is approximately 10 dB higher than for pure-tone stimuli (Figs. 8.14 and 8.15).
9.4 RESULTS: STATIONARY STIMULI

The results for the stationary stimuli are presented in this section. Only the results corresponding to S2 are shown, as the results for S1 are qualitatively similar.

9.4.1 Output profiles from the DRNL filter bank

The output profiles from the DRNL filter bank in response to the stationary stimulus S2 are shown in Fig. 9.4a. The stimulus spectrum is shown in Fig. 9.4b for comparison. (Note that Fig. 9.4a also shows the output profiles for unfiltered broadband noise stimuli of the same duration and rms intensity as the stimulus S2.)

At low signal levels [20 and 30 dB SPL (rms)] the broad spectral features (N1, P1, N2, P2 and N3) are still discernible in the DRNL filter profile. However, the fine spectral structure is lost (for instance, the rippling spectrum of the notch N2 disappears). Additionally, the maximum peak-to-notch amplitude (corresponding to P2/N3) is considerably smaller in the DRNL filter profile (around 13 dB) than in the stimulus spectrum (approximately 26 dB). These observations coincide with the conclusions made in Chapter 7 (see Introduction), which was somewhat expected because at low stimulus intensities the DRNL model behaves linearly (see Fig. 9.2a); that is, in exactly the same manner as the linear gamma-tone filter bank employed in Chapter 7. At low signal levels, therefore, the smearing of the spectrum can be only attributed to the energy spreading across the spectrum introduced by the filters' width, which increases with BF [see Eqs. (8.5) and (8.7)].

At moderate intensities [40 dB SPL (rms)], the output profile from the DRNL filter becomes even more smeared. For instance, the peak-to-notch amplitude of P2/N3 becomes around 8 dB. A 40 dB SPL stimulus is right in the middle of the compression region (see Fig. 9.2a). Therefore, one would expect two-tone suppression and distortion products to affect the spectral features of the stimulus spectrum maximally. Consequently, at moderate stimulus levels the smearing of the spectrum is likely to be caused by a double effect: (a) because of the energy spreading due to the filter bandwidths (i.e., the same effect as for signal levels in the linear
region of the filters), and (b) because of nonlinear-related phenomena as two-tone suppression and distortion products.

At higher stimulus intensities [60 dB SPL (rms)] even the broad spectral features are only vaguely discernible in the output profile from the DRNL filter bank. At 80 dB SPL (rms), the spectral features disappear almost completely and the profile decreases smoothly from around 50 dB at 4 kHz BF to 40 dB at 17 kHz. At these high stimulus intensities (particularly at 80 dB), the DRNL model becomes approximately linear again (see Fig. 9.3a). Therefore, nonlinear-related phenomena (e.g., two-tone suppression and distortion products) are unlikely to distort severely the spectral features of the stimulus. However, auditory filters become so broad\(^5\) (see, for instance, O’Mard and Meddis, 1996) at high signal levels that the energy spreading across BFs is very large. This reduces the spectral contrast in the output from the DRNL filter bank with respect to the original spectrum of the stimulus.

The gradual slope of the output profile from the DRNL filter bank at high stimulus intensities is a consequence of two factors: (a) the ear-canal resonance peak at 2.8 kHz (see Fig. 7.5) whose high-frequency skirt influences the response at 4 kHz, and (b) the attenuation, \(\alpha^W\), of the wide filter which becomes larger for higher BFs [see Eq. (8.8)].

An important conclusion can be already drawn from these results: pinna-based spectral features of the stimulus do not get past the cochlear filtering at very high stimulus intensities and, therefore, can not be reflected in the AN rate-profiles. If the spectral information is still present at high intensities, it must be present in some other form than simple AN rate profiles.

\(^5\) The explanation in terms of the DRNL filter bank (see section 8.2, Model description) is that at high stimulus intensities, the contribution of the narrow-filter signal to the total output from the DRNL filter is much smaller (because of the compression process) than that of the wide-filter signal, which becomes dominant. Therefore, the filter width increases at high stimulus intensities in the DRNL filter model. This reflects well-established physiological measurements (for a complete review see Pickles, 1988). Notice that the fact that the wide filter becomes dominant at high stimulus intensities is the reason the model becomes approximately linear at high intensities.
Figure 9.4. (a) Output profiles from the DRNL filter bank. The symbols represent the output profile for the stationary stimulus S2 (–20° elevation, 0° azimuth). The thick lines represent the output profile for pure broad-band noise. The rms stimulus intensity (dB SPL) is indicated by the numbers on the left of each line. Notice that the $\log_{10}(BF)$ is represented in the X-axis [the correspondence between the actual BF and the $\log_{10}(BF)$ is given in Fig. 9.1c]. (b) Spectrum of S2.
9.4.2 Onset and steady-state rate profiles associated with HSR fibres

As explained in Chapters 7 and 8, HSR fibres have the lowest rate thresholds (around 0 to 10 dB SPL) and a narrow dynamic range of around 30 dB. Therefore, it is interesting to investigate whether or not HSR fibres can convey the spectral information at high stimulus intensities. It has been reported that the quality of the rate representation is better at the stimulus onset, which has a wider dynamic range, than in the steady-state (Delgutte and Kiang, 1984a, b). We have used the model proposed in Chapter 8 to investigate this possibility further.

Figs. 9.5a and 9.5b show the onset and steady-state AN rate profiles for the stationary stimulus S2 at the five intensities considered. The figures show that the spectral features are clearly visible in both the onset and steady-state rate profiles at moderate stimulus intensities [30 and 40 dB SPL (rms)]. At lower intensities [20 dB SPL (rms)], the stimulus spectral shape is just discernible only in the steady-state rate profile (see also Fig. 9.6a). At 60 dB SPL (rms), the spectral features are only observed in the onset rate profile. Even in this case, however, the rate profile for BFs below around 6 kHz does not resemble the spectral shape of the stimulus. At 80 dB SPL (rms) neither the onset nor the steady-state rate profiles show any of the stimulus spectral features.

The AN rate profiles presented in Fig. 9.5 can be explained by looking at the output from the DRNL filter bank (Fig. 9.4) and the onset and steady-state RI for HSR fibres shown in Figs. 9.2b and 9.3a. For instance, the latter figures show that a 20 dB SPL (rms) stimulus is just below or at the rate threshold of HSR fibres. This explains why the broad spectral features are only just visible in the rate profiles for a 20 dB stimulus, even though they are clearly visible in the output profile from the DRNL filter bank (Fig. 9.4).

It is interesting to notice that the spectral contrast (i.e., peak-to-notch discharge rates) is higher in the onset rate profiles than in the steady-state ones. This must be a result of the larger discharge rates associated with the onset response compared with those in the steady-state (compare the y-axis scales of Figs. 9.2b and Fig. 9.3a). Another interesting observation is that the spectral contrast of the AN rate profiles (particularly
the onset one) for a 40 dB stimulus is very much larger than what might be expected from looking at the output profile from the DRNL filter bank. This is a direct consequence of the steep slope of HSR RI curves, which effectively increases the contrast of the AN rate profile with respect to the DRNL filter output profile.

9.4.3 Rate profiles of LSR, MSR and HSR fibres at different stimulus intensities

As explained in Chapter 7, LSR and MSR fibres are more likely to encode the spectral information for higher stimulus intensities because they have higher rate thresholds and wider dynamic ranges than HSR fibres (see for instance Fig. 9.3). In this section we examine the extent to which this is true by simulating AN rate profiles associated with the three types of fibres at various stimulus intensities. Fig. 9.6 shows the results of the model evaluation. The results shown in Fig. 9.6 correspond to the steady-state AN rate profiles in response to the stationary stimulus S2.

The visual analysis of the results yielded the following observations:

1. For a 20 dB SPL stimulus (Fig. 9.6a), the spectral features of the stimulus are only discernible in the rate profile of HSR fibres.
2. For a 30 dB SPL stimulus (Fig. 9.6b), the spectral features of the stimulus are visible in the rate profiles of the three types of fibres. However, for LSR fibres, the discharge rate is so small (around 4.8 spikes/s at the spectral peak P2) compared with that of MSR and HSR fibres that it is almost invisible in the figure.
3. For a 40 dB SPL stimulus (Fig. 9.6c), the spectral features of the stimulus are clearly visible in the rate profiles of the three types of fibres.
4. For a 60 dB SPL stimulus (Fig. 9.6d), the rate profile corresponding to HSR fibres is almost flat. The rate profiles of MSR and LSR fibres show a similar shape to the output profile from the DRNL filter bank (Fig. 9.4a). That is, the spectral features of the stimulus are visible in the rate profiles of LSR and MSR fibres for BFs above 8 kHz \(= 3.9 \log_{10}(\text{BF(Hz)})\). For lower BFs, the rate profiles of LSR and MSR show the decreasing slope toward high BFs observed in the output profile from the DRNL filter bank (Fig. 9.4a).
5. For a 80 dB SPL stimulus (Fig. 9.7e), the rate profiles of MSR and HSR fibres are almost "flat" and do not resemble the spectral shape of the stimulus.
Chapter 9: Simulating AN encoding of pinna-based spectral cues

The rate profile of LSR fibres shows the decreasing slope observed in the output profile from the DRNL filter bank (Fig. 9.4a).

These results can be interpreted and explained by looking at the RI curves for noise stimuli (Fig. 9.3) and the output profiles from the DRNL filter bank (Fig. 9.4). The most interesting result is, perhaps, that the degree to which spectral information is present in the AN rate profiles at high stimulus intensities is not only determined by the dynamic range of the AN fibres, but also, and most importantly, by the output from the cochlear filtering (i.e., from the DRNL filter bank in our model). This is evident, for instance, in the rate profile of LSR fibres for a stimulus intensity of 80 dB SPL (shown in Fig. 9.6e). In this case, the rate profile of the LSR fibre does not show the spectral features of the stimulus. However, the reason is not that the fibres are saturated at 80 dB SPL (as can be seen in Fig. 9.3c), but that the features are not present in the output profile from the DRNL filter bank (see Fig. 9.4a). In other words, if the spectral characteristics of the stimulus were present in the output profile from the DRNL filter bank for an 80 dB SPL stimulus, they would also be observed to some degree in the rate profile of LSR fibres, as the considered LSR fibres are not yet saturated at 80 dB SPL.
Figure 9.5. Simulated rate profiles for a HSR fibre in response to the stationary stimulus S2. (a) Onset response; calculated as the maximum response over the first 10 ms of the stimulus. (b) Steady-state response; calculated as the average response over last half of the stimulus duration. (See text for details).
Figure 9.6. Simulated AN rate profiles for LSR, MSR and HSR fibres in response to the stationary stimulus S2. (a) stimulus intensity 20 dB SPL (rms). (b) 30 dB SPL (rms); (c) 40 dB SPL (rms); (d) 60 dB SPL (rms); and (e) 80 dB SPL (rms).
9.5 RESULTS: MOVING STIMULUS

The model was evaluated for the moving stimulus MS. The results were analysed by looking at the maximum- and the average-rate profiles for each one of the two adjacent noise segments in the moving stimulus; that is, for the two HRTFs between which the source is switching (see Stimulus generation section above). Maximum-rate profiles for the first noise segment (NS1) represent the maximum discharge rate over the first 10 ms of the moving stimulus MS as a function of fibre BF. Maximum-rate profiles for the second noise segment NS2 represent the maximum discharge rates over the period 50-60 ms of the moving stimulus MS as a function of fibre BF. Notice that, since the moving stimulus was presented as a continuous 100-ms noise burst, the maximum-rate profiles for the first noise segment correspond to the onset rate profiles for the moving stimulus MS. The maximum-rate profiles for the second segment can be interpreted as an onset-like response as the stimulus switches from the first to the second location. Average-rate profiles for the two noise segments represent the average rates over the first and the last 50 ms of MS respectively as a function of fibre BF. The hypothesis behind our analysis is that there must be a noticeable change in the AN maximum-rate (or onset-like) response to changes in the stimulus location.

Figs. 9.7 and 9.8 show the maximum- and the average-rate profiles respectively for HSR fibres. Notice the resemblance between the rate profiles for the first noise segment (Figs. 9.7a and 9.8a) and the corresponding HRTF shown in Fig. 9.1a. As expected, the maximum-rate profiles for the first segment NS1 (Fig. 9.7a) correspond to the onset rate profiles for the moving stimulus MS. Maximum rates of around 1000 spikes/s are obtained at 80 dB SPL in that case. Maximum-rate profiles for the second segment NS2 (Fig. 9.7b) show rates around 400 spikes/s for an 80 dB SPL (rms) stimulus. However, larger discharge rates (around 770 spikes/s) are observed in the maximum-rate profiles for the second segment at much lower stimulus intensities (40 dB SPL rms). These large discharge rates occur only at BFs corresponding to the notch N1 (7.9 kHz) in the HRTF of the first source location (see Fig. 9.1a). In other words, the notch N1 in the HRTF of the first source location is clearly observed, in this particular case, as a peak in the maximum-rate profile for the second source location. Notice that this effect is maximal at 40 dB SPL (rms), but it is also visible at 30 dB SPL. However, it is not observed at lower or higher signal levels (20, 60 and
80 dB SPL rms). Additionally, the effect is much weaker in the average-rate profiles for the second source location (Fig. 9.8b) than in the maximum-rate ones.

This result can be interpreted as though the system is differentiating the HRTFs corresponding to two successive source locations, so that deep notches in the HRTF of the first location are transformed into clear peaks in the AN maximum-rate profiles as the source moves to a different location. Fig. 9.9 compares the maximum-rate (onset-like) profile for the second location (shown in Fig. 9.7b) and the difference between the HRTFs of the second and first source positions. It should be noticed, however, that this differential effect appears to be maximal only when there is a deep notch in the HRTF of the first source position followed by a spectral feature of larger amplitude (e.g., a peak or a flat spectrum) at the same frequency in the HRTF of the next source position; i.e., for a kind of increment-pedestal stimulus (Meddis, 1986).

The explanation to this interesting result is found in the processes that describe the flow of transmitter between the inner-hair cell (IHC) and the synaptic cleft (Meddis 1986, 1988). After a 50 ms signal, the transmitter contents in the free pool are in (or close to) equilibrium. When the stimulus intensity is suddenly increased (as it is the case for a moving stimulus such that a deep notch in the HRTF of one location is followed by a peak at the same frequency in the HRTF of the next location), the probability of a transmitter packet being released becomes suddenly larger (see Chapter 7, section 7.2.3: The mechanical to neural transduction). Therefore, more of the available packets in the free pool are released to the synaptic cleft causing the maximum discharge rate to increase momentarily, before reaching adaptation again (see figures 12 to 14 of Meddis, 1986). As shown in figure 14 of Meddis (1986), the effect is larger at the onset of the change, and it is maximal at moderate intensities, which explains our results.

According to Meddis (1986), a related effect must be observed for a decrement pedestal. In terms of moving stimuli, a decrement pedestal occurs, for instance, when spectral peak in the HRTF of the first location is followed by a notch in the HRTF of the second location. In this case, the notch of the second source location will be accentuated in the onset-like (or maximum-rate) profile for the second location. Unfortunately, this second effect is not so obvious in our moving stimulus because the notches in the HRTF of the second source location coincide approximately with notches in the HRTF of the first location. It is important to notice that this related
effect is also equivalent to differentiating two successive HRTFs. The result of the differentiation is noticed maximally in the maximum-rate (or onset-like) profile for the second location (Meddis, 1986).

A related result was reported by Poon and Brugge (1993a) after examining the response of individual AN fibres to notched-filtered broad-band noise. They reported that changes in individual AN fibres at BFs near spectral notches are large when the notch centre frequency changes (in an attempt to simulate switching from one HRTF to another). They also reported that the effect depends on the stimulus level; at signal levels above 70 dB, the presence of the notches was barely detectable in the firing pattern. Our results for the moving stimulus are consistent with the observations of Poon and Brugge. For instance, Fig. 9.7 shows that the change in discharge rate when switching from the first to the second HRTFs is maximal for the fibres at around 8 kHz BF (≈3.9 log_{10}(BF(Hz))). That is, at the position of the spectral notch N1 (Fig. 9.1) in the HRTF of the first location (S1).
Chapter 9: Simulating AN encoding of pinna-based spectral cues

Figure 9.7. (a) Maximum-rate (onset) profiles over the first 10 ms of the moving stimulus MS (i.e., for the first source location). (b) Maximum-rate (onset-like) profiles over the period 50-60 ms of the moving stimulus MS (i.e., as the stimulus switches from the first to the second location). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms).
Chapter 9: Simulating AN encoding of pinna-based spectral cues

Figure 9.8. (a) Average-discharge rate-profiles over the first 50 ms of the moving stimulus MS (i.e., for the first source location). (b) Average-discharge rate-profiles over the last 50 ms of the moving stimulus MS (i.e., for the second source location). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms).
Figure 9.9. (a) The ratio of the magnitude spectra of the two stimulus locations computed as $20 \log_{10} S2/S1$. (b) Maximum-rate (onset-like) profiles over the period 50-60 ms of the moving stimulus MS (i.e., as the stimulus switches from the first to the second HRTFs). The numbers on the right of each series correspond to the intensity of the moving stimulus expressed in dB SPL (rms). This figure is a reproduction of Fig. 9.7b.
Chapter 9: Simulating AN encoding of pinna-based spectral cues

9.6 DISCUSSION

The computer model of the peripheral auditory system presented in Chapter 8 has been employed to simulate AN rate profiles to broad-band noise stimuli filtered through human-like HRTFs. Two types of stimuli have been used: stationary stimuli and moving stimuli. The output profiles from the DRNL filter bank have been examined for the stationary stimuli at different intensities. The quality of the onset and the steady-state AN rate profiles for HSR fibres has been compared. The quality of the steady-state rate profiles for LSR, MSR and HSR fibres has been examined at various stimulus intensities. Finally, the model has been evaluated for a moving stimulus. The model results for this stimulus have been presented in the form of maximum-rate (or onset) and average-rate profiles for each one of the two successive source locations.

9.6.1 Cochlear output profiles

An important finding of this study is that even the broadest spectral features of the stimulus are not observed in the output profiles from the DRNL filter model at high stimulus intensities (Fig. 9.4). As explained above, this must be primarily a consequence of the large filter widths observed at high stimulus intensities. It must be pointed out that this result is a prediction of the DRNL filter bank model and is subject to empirical testing. So far as we know, there are no experimental reports on the simultaneous response to broad-band stimuli of a large number of points (i.e., frequency channels) along the BM. If the model's prediction is confirmed, it would mean that the location-dependent spectral information available in HRTFs is not transmitted through the mechanical cochlear filtering at high stimulus intensities.

It is known, however, that the formant bands of vowel sounds are only evident in experimental AN rate profiles for vowels presented at low levels (Sachs and Young, 1979; Young and Sachs, 1979). However, the formant bands in the rate profiles disappear if the vowel level is increased. Generally, this phenomenon has been attributed to the rate saturation of the AN fibres. Because the dynamic range of the majority of AN fibres (HSR fibres) is limited to approximately 30 dB, several mechanisms have been proposed which the auditory system could use to encode the
spectral information in the AN rate response at high stimulus intensities. In this sense, it has been proposed that for high stimulus intensities the spectral characteristics of the stimulus are much better encoded at the stimulus onset than in the steady-state because the onset response has wider dynamic range (Delgutte and Kiang, 1984a,b). Another possibility (that has been examined in this chapter and previously in Chapter 7) is that LSR fibres are more appropriate to encode the spectral information at high stimulus intensities because they have higher thresholds and wide dynamic ranges. These two possibilities have been explored here. The results (Figs. 9.5 and 9.6) have shown that the two proposed mechanisms may be actually contributing to the AN encoding of the stimulus spectrum at moderately high (around 60 dB SPL rms) stimulus intensities. However, it has been found that the degree to which the spectral features of the stimulus are encoded in the AN rate profiles at higher stimulus intensities (80 dB SPL rms) is primarily determined by the output profile from the cochlear filtering. Therefore, even though the onset dynamic range and/or the dynamic range of LSR fibres may be wide enough to encode the spectral features of the stimulus at high stimulus intensities, those features are not observed in the BM motion, which, after all, constitutes the actual driving force for AN excitation.

The implications of this result are very important from a psychophysical point of view, as it suggests that sound localisation must decrease with increasing stimulus level in tasks where only spectral cues are available. A related psychophysical result has been reported by Hartmann and Rakerd (1993). They concluded that sound localisation accuracy for click stimuli in the median sagittal plane decreases with increasing stimulus intensity. They attributed this effect to the fact that the spectral peaks and valleys, which normally code for localisation in the median plane, are less recognisable because of the broadening of the tonotopic excitation pattern at high signal levels. Our results agree, therefore, with Hartmann and Rakerd's conclusions.

---

6 Hartmann and Rakerd (1993) actually found that this is true only for impulsive sounds (clicks) but not for broad-band noise stimuli. The results of their localisation experiments and discrimination experiments showed that "...the spectral details are perceived better for noise than for clicks". However, they argue that one possible explanation to this discrepancy may be related to the role of lateral inhibition in re-sharpening the spectral characteristics of the stimulus. They argue that "...if the operation of inhibition is delayed compared to excitation (which broadens with increasing signal level), then the (spectral) details of noise should be resolved better that the details of intense clicks, in agreement with experiment". Further research is required to confirm or reject this explanation.
9.6.2 AN rate profiles for stationary stimuli

With regard to the AN rate profiles, the results of the model show that the spectral characteristics of the stimulus are directly observed to some degree in the rate profiles of at least one type of AN fibre for signal levels below 80 dB SPL (rms). The spectral information of the stimulus is optimally encoded in the rate profiles of HSR, MSR and LSR fibres for low, medium and moderately-high (60 dB SPL) signal levels respectively. These results should be comparable with the experimental observations of Rice et al. (1995), as the evaluation paradigms employed in both studies are almost identical. At first sight, however, our results appear to disagree with their observations, as they found very scattered rate profiles, thus very little resemblance between the unprocessed AN rate profiles and the stimulus spectrum. Rice et al. (1995) concluded that the variability in the rate characteristics of AN fibres explains this result (fibres differ in spontaneous rate, in saturation rate, and in the slopes with which their rates increase with stimulus level). They also concluded that this variability is eliminated in the rate-difference profiles between the rate profiles for two stationary stimuli at different source locations. In other words, by using each fibre as its own control.

Our model does not include the experimental variability amongst fibres specifically, which explains why, in most cases, the spectral features of the stimulus are clearly observed in the simulated rate profiles (see Discussion in Chapter 7). As a matter of fact, some unintentional variability exists across BFs in the saturation rates and in the slopes of the RI curves. This variability is reflected in the fact that the RI curves are not identical at 4 and 16 kHz BF (see Figs. 9.2 and 9.3) and is a consequence of using the same value for the parameter $h$ across BFs (see Chapter 8 and Table 9.1). In any case, the slight variability included in the model is not enough to account realistically for the experimental differences amongst fibres. In order to simulate the results by Rice et al. (1995), specific fibre variability should be included in our model by allowing some of the model parameters (perhaps $h$ and $z$) to vary 'randomly' within reasonable values.

---

7 There is one case amongst our results in which the broad spectral characteristics of the stimulus are clearly observed in the rate profiles of the three fibre types. This case corresponds to a 40 dB SPL stimulus (Fig. 9.6c).
9.6.3 Moving Stimuli

The model has been also used to examine the AN rate response to stimuli moving between two locations with very different HRTFs. It has been found that AN maximum-rate profiles calculated within a short interval after the source switches from one location to another, resemble the first order differential of the HRTFs for the two locations. The effect is maximal at moderate signal levels (30-40 dB SPL) within the dynamic range of the fibre.

We can speculate about the practical applications of this mechanism by the auditory system. One possibility is that this mechanism may aid in the detection of spectral notches. Moore et al. (1989) provided evidence that subjects were capable of detecting spectral stationary peaks and notches at least at 1 and 8 kHz, and that the sensitivity to peaks was greater than to notches. Additionally, they found that changes in the centre frequency of the notch could be detected more easily than stationary notches. The differential mechanism proposed above would bring together these two results under a single rationale, namely the sensitivity to peaks (of AN activity) is greater than to notches. We have shown that in some cases (see above), spectral notches observed in the HRTF for a given source location are seen as spectral peaks in the AN maximum-rate profile when the sound source switches to a different location. In other words, in some cases, moving notches are reflected as peaks in the AN rate response, which provides a possible explanation to the fact that moving notches are more easily detectable than stationary ones (Moore et al., 1989).

In addition to this, there is evidence that supports the possibility that a serial differencing mechanism may be used to aid spectral estimation (a revision of the evidence is provided by Rice et al., 1995). For instance, a differencing mechanism has been proposed by Rice et al. (1995) as a possible way of eliminating the experimental fibre variability which prevents the spectral characteristics of the stimulus from being encoded in simple rate profiles. The differencing mechanism for moving sources predicted by the model constitutes a possible candidate. However, it has been shown above that this mechanism is restricted to a very short range of signal levels at least for HSR fibres. Assuming that the differencing properties of the IHC/AN synapse are directly related to the dynamic range of the fibres, it may be that for MSR
and LSR fibres the proposed mechanism is operational over a wider range of stimulus intensities. The issue requires further investigation.

One important characteristic about the differencing mechanism predicted by the model is the fact that it occurs at a very early stage in the auditory system, namely at the IHC/AN synapse. In other words, it does not require any neural processing circuitry at higher stages in the central nervous system. This is not to say, however, that this circuitry and/or other serial differencing mechanisms do not exist at higher levels in the central nervous system.

9.7 SUMMARY AND CONCLUSIONS

We have used the nonlinear auditory model presented in Chapter 8 to simulate the AN rate response to stationary and moving broad-band noise stimuli filtered through human-like HRTFs. The main findings and conclusions are:

1. The degree to which the spectral characteristics of the stimulus are preserved in the AN discharge patterns at high signal levels is primarily determined by the tonotopic representation of the stimulus spectrum at the cochlea, rather than by the dynamic range of AN fibres. In this sense and with regard to the sound localisation problem, it has been found that the spectral characteristics of HRTFs are not preserved in AN rate profiles at high signal levels because of the broadening of the cochlear filters. This suggests that sound localisation accuracy must decrease with increasing stimulus level in tasks where only HRTF spectral cues are available.

2. Transient AN maximum-rate profiles calculated as the stimulus switches from one location to another resemble (to a first approximation) the first order differential of the HRTFs associated with the two successive locations. It has been speculated that this mechanism may aid in the detection of moving spectral notches. Additionally, it has been explained that a differencing mechanism of this kind may be used to eliminate the experimental variability in the rate characteristics of AN fibres, which would aid in spectral estimation.
tasks. The main limitation of the described differencing mechanism is that it is only functional at moderate signal levels, at least for HSR fibres.

3. Much research has yet to be done on the mechanisms by which spectral information is signalled to the brain. Mechanisms for spectral estimation different from those proposed above may be at work at higher nuclei in the central nervous system. If these mechanisms exist, they are yet to be discovered. In any case, much progress can be made by using computer modelling to investigate the issue now that an auditory model which includes the nonlinear filtering by the BM is available.
CHAPTER 10

General discussion and conclusions

10.1 SUMMARY

The work presented in this thesis has provided some insights into the physical origin and the physiological coding of pinna based spectral cues. An experimental technique for measuring head-related transfer functions has been presented. This technique has been employed for characterising the universality and the source-location dependency of the spectral features observed in HRTFs, particularly those related to the pinna. A physical diffraction/reflection model of the concha cavity has been developed and tested. This model has been employed for investigating a theoretical explanation to some of the most remarkable spectral notches observed in experimental HRTFs and their dependency on source location. The contribution of other anatomical features of the pinna (fossa, scapha, pinna-flange, tragus and crus helias) to the overall HRTF has been measured experimentally.

The physiological coding of pinna-based spectral cues has been examined with two models of the auditory system. Firstly, a model that includes a linear gamma-tone auditory filter bank has been employed. Secondly, a more exhaustive model that includes a novel dual-resonance nonlinear (DRNL) auditory filter bank has been used. A great effort has been devoted to developing, perfecting and optimising this latter model so that it embodies the current knowledge on the auditory system. The response of both models to stationary HRTF-filtered broad-band stimuli has been presented, and conclusions have been drawn about the effect that the rate characteristics of auditory nerve (AN) fibres has on the rate representation of pinna-based spectral cues. Additionally, the nonlinear model has been used to examine the transient AN rate representation to moving stimuli.
This chapter is concerned with summarising and discussing the major findings and limitations of the research activities listed above that are covered in this thesis. The main conclusions drawn from this research are presented (in italics) and ideas for future research projects are proposed.

10.2 DISCUSSION and CONCLUSIONS

10.2.1 Technical considerations: An experimental method for making HRTF measurements using a KEMAR fitted with individualised moulded pinnae

Pinna-moulding technique

A technique for making individualised-moulded pinna to be fitted into a KEMAR has been documented. Although KEMAR manufacturers provide a diversity of ears to be employed with the manikin, the range of ears is limited to a few models. Furthermore, KEMAR ears do not correspond to real ears; but rather they are designed to provide an average response (Burkhard and Sachs, 1975). For these reasons, some researchers have expressed their scepticism about the generality of HRTF measurements obtained with KEMAR ears, on the grounds that differences must exist between them and the ears of real people. *The method for pinna moulding documented in this thesis alleviates this problem to the extent that it allows pinna casts of real ears to be fitted into the KEMAR*. The method has been validated by comparing our HRTF results for three pairs of individualised-moulded pinnae with HRTF data for real ears previously published. Overall, the shape of the HRTF coincide in both cases. Differences between our HRTFs and the HRTF for real ears are attributed to expected differences in pinna shapes. However, in order for the pinna-moulding method to be properly validated, a comparison should be made between HRTFs obtained with real subjects [using, for instance, the technique by Pralong and Carlile (1994)] and those obtained with a KEMAR wearing moulded pinnae from the same subjects.
An experimental method for measuring HRTFs

An experimental method for measuring head-related transfer functions (HRTFs) using a KEMAR was already in use in our laboratory prior to the beginning of this research project. (The method has been fully documented in Chapter 3). However, the technique has been improved considerably during the course of the research covered in this thesis (the improved technique is described in Chapter 6). New software has been written so that HRTF measurements can be performed in real time using the audio facilities of an Indy Silicon Graphics work-station. The new technique requires considerably less time for each HRTF measurement (see Chapter 6). Additionally, and most importantly, the new software produces reliable HRTF measurements for frequencies above around 2-3 kHz even in reverberant rooms. The performance of this improved method in anechoic (or free-field) conditions has not been fully investigated. However, it seems reasonable to think that in those conditions the only limiting factors will be associated with the quality of the hi-fi equipment.

10.2.2 The location dependency of pinna-based spectral features

An important part of this thesis has been dedicated to investigating the generality of pinna-based spectral features (peaks and notches) present in HRTFs, and to characterise their dependence on source location. Extensive work on this area has been previously done by Shaw and Teranishi (1968), Teranishi and Shaw (1968) and Shaw (1975). However, their work focused mainly on spectral peaks and, in this respect, our work has mainly corroborated the previous knowledge.

With regard to spectral notches, however, the contribution of this thesis is considerable. Our work was originally motivated by the existence of an unexplained elevation-dependent spectral notch (N1) whose centre frequency increases from around 6-7 kHz to 10-12 kHz, as the elevation angle of the sound source is increased from −40° to +40°. This notch had been observed for sources in the median vertical plane (Hebrank and Wright, 1974; Butler and Belendiuk, 1977) and also in the lateral vertical plane (Shaw and Teranishi, 1968; Bloom, 1977). In this thesis, the location dependency of this notch has been characterised more precisely. It has been shown that this elevation-dependent notch is consistently observed not only in the median and lateral vertical planes but for all ipsilateral locations in the frontal quadrant, and even to
some extent (see Chapter 4), for contralateral locations. The azimuthal dependency of the notch has also been examined at a constant elevation (Chapters 3 and 4). It has been shown that the centre frequency of the notch is virtually invariant as a function of source azimuth (its centre frequency is only around 1 kHz higher in the lateral vertical plane than in the median plane); in many cases, this dependency is almost unnoticed.

It has been suggested that it is physically possible that such an elevation-dependent notch is responsible for the sensation of elevation for any source position within the frontal hemi-field. This suggestion is supported by (1) the psychophysical evidence that an elevation-dependent notch encodes source-elevation information for sources in the median plane (Hebrank and Wright, 1974; Butler and Belendiuk, 1977) and in the lateral vertical plane (Bloom, 1977), (2) by the physical evidence presented in this thesis that such a notch is also present for sources in intermediate planes between these two and, in some cases, for contralateral sources, and (3) by the fact that the centre frequency of the notch is largely invariant as a function of azimuth, as shown in this thesis (Chapters 3 and 4). In any case, this evidence is only circumstantial and further psychophysical studies should be carried out to investigate our suggestion formally.

The HRTF data presented in this thesis have revealed two other major notches at higher frequencies (N2 at around 10-12 kHz, and N3 at around 15-17 kHz, for a source at −40° elevation). N2 is only observed at low elevations. However, N3 seems to depend on source elevation in a similar way to N1 (described above). Both notches, N2 and N3, show a slight dependence on azimuth (Chapters 2 and 3). This dependence is more evident for N3 as it has higher centre frequency — overall, there is a tendency for spectral features at higher frequencies to show a stronger dependence on azimuth. The similarities in the azimuthal and elevation dependency of N1 and N3 indicate that they may be generated by the same physical process.

10.2.3 Modelling the transfer function of the pinna: The physical origin of pinna-based spectral features

A major contribution of this thesis is a physical model of sound diffraction, reflections and interference in the human concha (Chapter 5). Such a model has proven to be very useful for explaining how the dependence of pinna-based spectral notches comes about.
Prior to this work, the physical processes that generate the characteristic elevation-dependent spectral notches described above were unknown (see Chapter 5). The suggestion by Hebrank and Wright (1974) that the elevation-dependent notches may be caused by cancellation between the direct wave entering the meatus and the reflected wave from the concha posterior wall was interesting but their single-delay-and-add formulation was too simplistic. Although the diffraction/reflection model developed in this thesis is inspired by the idea of concha reflections suggested by Hebrank and Wright (1974), its formulation is considerably more realistic and accurate than the single-delay-and-add model. As a matter of fact, the diffraction/reflection model is an approximate solution to the wave equation for an acoustic system such as the concha. Therefore, diffraction (scattering) of the sound as it enters the concha cavity and in the reflection process, and reflections from a large number of points on the posterior wall of the concha are naturally included in the formulation. To our knowledge, the diffraction/reflection model constitutes the first successful attempt to simulate the transfer function of the concha from first physical principles.

The diffraction/reflection model predicts two concha-related spectral notches which have similar centre frequencies and similar dependence on azimuth and elevation as N1 and N3 (described above). According to the diffraction/reflection model, N3 corresponds to the $3f_{N1}$ minimum associated with N1. Additionally, the diffraction/reflection model suggests that the elevation dependency of N1 and N3 comes about as a result of the asymmetrical geometry of the posterior wall of the concha with respect to the meatus entrance. In this sense, Hebrank and Wright's idea that the elevation dependency of N1 comes about because different source elevations have different concha delay-paths associated is partially correct. However, it has been shown that multiple concha reflections (or delays) properly weighted in amplitude, must be considered in order to predict the centre frequency of spectral minima accurately. Finally, according to the diffraction/reflection model, the scattering (diffraction) of the sound as it enters the concha cavity explains why similar elevation-dependent spectral notches, N1 and N3, are observed for sources in any vertical plane within the frontal hemi-field. The explanation is as follows: because the incident sound is scattered as it enters the concha cavity, reflections on the concha posterior wall occur consistently and similarly at any azimuthal angle. Therefore, interference between the direct wave entering the meatus and the reflected waves from the concha.
posterior wall generates spectral notches N1 and N3 whose centre frequency is predominantly dependent on elevation.

The main criticism of the diffraction/reflection model is the number of approximations and assumptions (physical and geometrical) involved in its formulation and evaluation. As a result, the diffraction/reflection model must be considered as an approximation and is open to improvement. Additionally, the diffraction/reflection model has been restricted to the concha cavity, despite the evidence presented in this thesis (Chapter 6) that other anatomical features of the pinna (the flange cavities, for instance) also contribute substantially to the total transfer function of the pinna. These other anatomical features should be included in an improved version of the diffraction/reflection model.

With regard to producing a complete model of the pinna transfer function, a review of the facts presented in this thesis shows that modelling the transfer function of the pinna can not be done by modelling and 'adding' the individual transfer functions of particular anatomical features of the external ear. This conclusion is supported by the results presented in Chapter 6, where it is also shown that it is not possible to establish a one-to-one relationship between specific pinna-based spectral features and individual anatomical features of the pinna. For instance, it has been shown that N1 is largely influenced by the occlusion of the flange cavities (fossa and scapha), despite the strong evidence that it is actually generated in the concha. This suggests that a functional model of the transfer function of the pinna can only be achieved by modelling the full pinna as a single system, so that the effects of acoustic interactions between the different anatomical features are reflected in the overall transfer function.

Some ideas as to how to find an exact physical description of the pinna are given below in the section Proposed future work.

10.2.4 A computer model of the auditory peripheral system

An important contribution of this thesis is in the form of a novel, exhaustive computer model of the peripheral auditory system (Chapter 8). In this thesis, the development

---

1 'adding' when the individual transfer functions are expressed in dB, or 'multiplying' when they are expressed in pressure units.
of such a model was motivated by the interest in investigating the physiological coding of pinna-based spectral cues at the auditory nerve (AN) level (this issue is summarised and discussed in the next section). However, the proposed model is a general one and, therefore, can serve as a firm platform for future investigations about higher nuclei in the auditory system.

The proposed model brings together some existing models of specific processes of the peripheral auditory system [such as the dual-resonance nonlinear auditory (DRNL) filter bank by O'Mard and Meddis (1996), the inner hair-cell (IHC) model by Shamma et al. (1986), and the synaptic-effects model by Meddis (1986, 1988)] with new ones, such as the transmitter-release exponential function. In this sense, the main contribution of this thesis is not to develop individual models of these specific peripheral processes, but to improve them and combine them in order to provide a global, exhaustive and robust model which embodies the current understanding of the peripheral auditory system. Special efforts have been given to generate an optimum set of parameters for each model stage so that the model simulates corresponding experimental data (e.g., cochlear I/O functions, two-tone suppression, distortion products, IHC AC/DC and DC-I curves, and AN RI curves for LSR, HSR and MSR fibres). In some cases, it has been necessary to change the parameter values with respect to those originally proposed by the author(s) of the corresponding model (e.g., the IHC model by Shamma et al., 1986).

At the AN level, the resultant model simulates the rate activity of low-, medium- and high-spontaneous rate AN fibres (LSR, MSR and HSR respectively) for stimulus frequencies at and off the fibre's best frequency (BF), for BFs across the range 4 kHz to 18 kHz. This is possible thanks to the incorporation of two key features in the model: (a) the DRNL auditory filter bank, and (b) the new transmitter-release exponential function. With regard to the DRNL filter bank, it determines the off-BF response of the simulated AN fibres. Therefore, the agreement between experimental and simulated off-BF RI curves (Chapter 8) is a direct consequence of the adequacy of the DRNL auditory filter-bank.

As for the transmitter-release exponential function, it allows the model to reflect two important aspects of the IHC/AN synapse. Firstly, all three fibre types (LSR, MSR and HSR) are driven by the same IHC intracellular potential, which agrees with the experimental evidence provided by Liberman and Oliver (1984) that AN fibres of the
three types are known to be in contact with the same IHC. Secondly, *the physiological differences between LSR, MSR and HSR fibres are presynaptic, in the relation between the transmembrane potential and the transmitter release*, which agrees with the experimental evidence provided by Liberman *et al.* (1990) that there exist presynaptic morphological differences for LSR, MSR and HSR fibres. It has been suggested in this thesis that the parameters $h$ and $z$ of the transmitter-release exponential function may be related to the dimensions the synaptic body and plaque respectively. Further research needs to be carried out about the actual function of these presynaptic structures in order to confirm or reject this suggestion.

Although the proposed model serves its purpose, it is certainly open to improvement. Because we were only interested in using this model for investigating the AN rate-representation of pinna-based spectral cues which occur above 4 kHz, the model has been only optimised to simulate AN RI curves for BFs between 4 and 18 kHz. Therefore, the validity of the model at BFs below 4 kHz needs to be tested. Additionally, the adaptation characteristics of the simulated AN fibres have only been examined visually in order to make sure that reasonable responses were obtained. Although that has been always the case, a quantitative investigation about the simulated adaptation time-constants should be carried out.

10.2.5 The physiological coding of pinna-based spectral cues

*Stationary stimuli*

The quality of the AN rate-representation of pinna-based spectral features has been investigated by modelling the AN response to HRTF-filtered broad-band stationary stimuli. Two models of the peripheral auditory system have been used; a model that includes a linear gamma-tone filter bank (Chapter 7), and the more realistic model described right above, that includes the DRNL filter bank. The interest behind using the two models was to examine to what extent cochlear nonlinearities determine the quality of the AN rate-representation of pinna-based spectral features at different signal levels.

Several conclusions have been drawn. (1) The fine spectral structure of the stimulus spectrum is lost in its corresponding AN rate profile; that is, only broad spectral
features are signalled to the central nervous system. (2) The quality of the AN rate-representation of the stimulus spectrum gradually deteriorates as a function of signal level, so that at around 80 dB SPL, not even the broadest spectral features are observed in the corresponding AN rate profiles. (3) The psychophysical implications of these findings are clear: only broad spectral features are available as localisation cues and sound localisation accuracy should deteriorate at high signal levels in tasks where only pinna-based spectral cues are available. Experimental psychophysical evidence has been presented to support that this is actually the case.

Prior to this research, it was already known that the quality of the AN rate profiles deteriorates as the intensity of the stimulus is increased (Sachs and Young, 1979; Rice et al. 1995). The same effect is seen in the simulated rate profiles for the nonlinear model (Chapter 9), but not for those obtained with the linear model (Chapter 7). Sachs and Young attributed this deterioration mainly to two factors: (1) two-tone suppression phenomena as a result of the nonlinear cochlear filtering, and (2) the limited dynamic range of AN fibres. These hypotheses have been amply tested with both models and a careful analysis of the results suggests that the broadening of the cochlear filters as a function of stimulus intensity is the factor mostly responsible for the deterioration of the AN rate profiles at high signal levels. It has been also shown that this factor determines the quality of the rate profiles both at the stimulus onset and in the steady-state. In this sense, the model shows that the broad spectral features of the stimulus are more clearly seen in the onset rate-profiles than in those associated with the steady-state [as suggested by Delgutte and Kiang, (1984a,b) for speech signals], but only for signal levels at which those features are ‘visible’ in the output profiles from the cochlear filtering (Chapter 9).

Moving stimuli

The nonlinear model of the auditory peripheral system has been employed to investigate the transient AN rate profile of the stimulus spectrum as it switches from one HRTF to another (i.e., moving stimulus). To our knowledge, this is the first study (experimental or by modelling) of this kind. Therefore, the results obtained and the conclusions made from this investigation are subject to experimental confirmation.
An important conclusion of this investigation is that for signal levels within the dynamic range of the fibre, the transient response between two locations (i.e., between two HRTFs) resembles the first order differential of the two corresponding HRTFs. It has been shown that in some cases this causes moving spectral notches (e.g., the elevation-dependent spectral notches described above) to be reflected as 'peaks' in the transient AN rate profiles. It has been suggested that this explains the fact that moving spectral notches are more easily detectable than stationary ones (Moore et al., 1989) and may provide some basis for a model of sound localisation based solely in the detection of peaks of AN activity by the higher nuclei in the auditory system.

10.2.6 The ambiguity of spectral cues in sound localisation: What happens if the sound source does not have a flat spectrum?

This thesis would be incomplete without a brief discussion on the issue of ambiguous spectral cues. Any model of sound localisation based on spectral cues must acknowledge the importance of a priori information regarding the source spectrum. In Chapter 2 (Background), it was explained that the signal at the eardrum can be expressed as the product: \( E(j) = A(j)S(j) \), where \( A(j) \) is the spectrum of the signal at the eardrum, \( A(j) \) is the head-related transfer function, and \( S(j) \) is the source spectrum. In other words, the presence of a particular feature at the eardrum can correspond to a location-dependent spectral feature present in the HRTF, or it could be a feature that was present in the source spectrum. The experiments conducted by Blauert (1969/70), Hebrank and Wright (1974), Bloom (1977), Watkins (1978) and many others (for a review see Middlebrooks and Green, 1991) indicate that the auditory system cannot effectively distinguish between the two — in all these experiments, the manipulation of the spectral contents of the stimulus presented at the eardrum produced source-location illusions.

Despite this evidence, mechanisms are being constantly suggested which may be used by the auditory system to extract the location-dependent spectral information, \( A(j) \), from the compound signal at the eardrum, \( E(j) \). One such mechanism was proposed by Searle et al., (1975, 1976) and consists in comparing the signals at left and the right ears to eliminate ambiguity introduced by the source spectrum. Rice et al. (1995) suggested a serial (thus, monaural) time-differencing mechanism so that localisation estimation is aided by finding the difference between the current responses and the
responses in the immediate past. This serial differencing mechanism also has the effect of eliminating the dependence on the source spectrum [as well as eliminating the noisy representation related to the variability in the discharge characteristics of AN fibres (see Chapter 9)]. Zakarauskas and Cynader (1993) proposed a theory for extracting $A(f)$ from the ambiguous signal at the eardrum, $E(f)$, by finding the second differential with respect to the frequency (i.e., between adjacent frequency channels). The main condition for the success of this mechanism is that $S(f)$ has to be smooth (but not necessarily flat) and broad-band.

Whether or not these theoretical mechanisms are actually available to (and/or used by) the auditory system is dependent upon the existence of the appropriate neural circuitry for them to be carried out. The issue is widely open to investigation.

10.3 PROPOSED FUTURE WORK

An attempt has been made in this thesis to shed some light on the physical processes underlying the generation of pinna-based spectral cues and the issues involved in the physiological coding of those cues by the peripheral auditory system. Although progress has been made, the investigation is by no means complete. There remain, as ever, many puzzling questions and problems yet to be solved. It is hoped, however, that this thesis lays the groundwork for future research in this field.

In the course of this research some suggestions for future research projects have been proposed. Here is a summary of some of the main ideas:

1. HRTF recordings: to compare HRTFs measured from real subjects with those for a KEMAR wearing moulded pinnae from the same subjects. [A technique for measuring HRTFs using real subjects has been provided by Pralong and Carlile (1994)]. If the comparison is successful, this will provide a general method for making individualised HRTFs measurements without the need for real subjects to suffer long recording sessions.

2. Psychophysical experiments to investigate whether the elevation-dependent notch N1 is actually responsible for the sensation of source elevation for sources in all
vertical planes within the frontal hemi-field. The experimental paradigms employed by Hebrank and Wright (1974) or Bloom (1977) could be used. The influence of the depth and width of the notch could be tested by including them as parameters in the experimental design. Such an investigation would provide stronger evidence of the role of elevation-varying spectral notches not only as cues for localisation in the median sagittal plane, but as general spectral cues responsible for the sensation of elevation for sources anywhere in the frontal acoustic space.

3. To find an exact physical model of the transfer function of the pinna. An exact physical model of the transfer function of the pinna could in theory be developed by solving numerically (or by a finite analysis method) the wave equation (see Appendix 2) for a particular pinna geometry. Such a solution [Eq. (A2.1)] can only be obtained if a precise geometrical description of the pinna is available, so that geometrical approximations are not necessary. Such a description could be obtained by *digitising the pinna surfaces* with a high spatial sampling-frequency (i.e., a large number of points). Note that the exact solution of the wave equation [i.e., Eq. (A2.1)] can be only found for a closed surface and, therefore, a geometrical description of the inner and the outer surfaces of the pinna would be required. An exact model of the acoustic transfer function of the pinna also requires the acoustic impedance of the pinna walls to be precisely known. Once the pinna geometry and its acoustic impedance are measured, the numerical evaluation of Eq. (A2.1) can be done (non-trivially) as proposed by Filippi and Dumery (1969) and Cassot (1975) for the case of thin totally-reflecting screens.

An exact model of the transfer function of the pinna could be used to investigate the precise origin of the location dependency of pinna-based spectral features. Additionally, it could be used for producing computer-generated individualised HRTFs, which may be of interest in virtual reality research and applications. Potentially, it may also be applied in the development of more sophisticated hearing aids.

4. To improve the computer model of the peripheral auditory system. As explained above, the functionality of the proposed model is restricted to BFs above 4 kHz. Further research should be carried out intended for extending this functionality to lower BFs, as well as for studying the adequacy of the simulated adaptation time-constants. The result would be a complete, functional model of the auditory peripheral
system across the whole frequency range. Such a model could be used as a solid platform for further auditory research both by modelling and experimentation. In some cases, such a model may be used as a substitute for pilot, redundant or unnecessary experimental studies with animals. In other cases, it may provide suggestions for more substantial and innovative experimental studies.

5. Transmitter release at the IHC/AN synapse. Further research should be carried out in order to establish precisely the role of the synaptic body and plaque at the IHC/AN synapse. This would provide accurate information about the relationship between the transmitter release rate and the IHC transmembrane potential which will confirm or reject the adequacy of the transmitter-release exponential function proposed in this thesis.

6. To investigate plausible mechanisms for extracting spectral information at high signal levels. Further psychophysical research should be carried out to investigate whether sound localisation accuracy actually deteriorates as a function of signal level in localisation tasks where only spectral cues are available. If that proves not to be the case, possible mechanisms employed by the auditory system to extract location-dependent spectral information at high signal levels must be investigated since we have seen that simple rate profiles do not convey such information. The computer model of the peripheral auditory system provides an interesting alternative to physiological experimentation for further investigation in this field.

7. To investigate plausible mechanisms for resolving spectral ambiguity. The problem of spectral ambiguity has been explained above. Some differential mechanisms have been presented as a solution. Further research should be done to investigate whether the neural circuitry necessary for implementing those (or other) differential mechanisms actually exists in the auditory system.
References


LUTEar (1993). LUTEar is a flexible computing tool for auditory simulation. It runs on a variety of platforms, and is available over the internet via anonymous FTP. ftp://suna.lboro.ac.uk/public/hulpo/lutear/www/linklutear1.html


References


References


APPENDIX 1

Pinnae photographs

Photographs of the pinnae employed for measuring the head-related transfer functions (HRTFs) presented in Chapters 3 and 4 are shown in this appendix. All the photographs shown correspond to pinna casts designed to be fitted onto a KEMAR.

Two types of pinnae are shown:

(1) Standard pinnae provided by Knowles (i.e., the KEMAR manufacturers). Two different pinna models are shown DB-061 (small, pink, left pinna) and DB-065 (large, red, right pinna). The HRTF data presented in Chapter 4 were obtained with these pinnae.

(2) Individualised moulded-pinna obtained with the pinna-moulding technique described in Chapter 3. Three pinna models are shown: IX, AC and LPO. The HRTF data presented in Chapter 3 were obtained with these pinnae.

Additionally, a photograph of the infills or 'flush-fittings' is also shown. Notice that the infills are simply a pair of casts which produce the effect of having no pinnae (i.e., the effect of having a pair of holes on both sides of the head) when fitted onto the KEMAR.
Appendix 1: Pinnae photographs

Knowles-manufactured KEMAR pinna

DB-061

Knowles-manufactured KEMAR pinna

DB-065
Individualised moulded pinnae for subjects IX (top) and AC (bottom)
Individualised moulded pinnae for subject LPO (top), and Infills (bottom)
APPENDIX 2

Mathematical development of the diffraction/reflection model

The development of the mathematical formulation of the diffraction/reflection model (presented in Chapter 5) is explained in this Appendix. General diffraction theory is only briefly introduced. For a complete review on diffraction theory see Hecht (1990).

A2.1 General

A general solution for the wave equation in a bounded or unbounded medium with no sound sources is given by (Morse and Ingard, 1968):

$$p(r) = \int_S \left[ g(\omega, R_o) \nabla p(r_o) - p(r_o) \nabla g(\omega, R_o) \right] dS_o$$  \hspace{1cm} (A2.1)

where $R_o = |\overrightarrow{r} - \overrightarrow{r_o}|$ and the integration is done over any closed surface, $S$, which surrounds the observation point $\overrightarrow{r}$. $g(\omega, R_o)$ is the Green's function whose value is:

$$g(\omega, R_o) = \frac{\exp(jkR_o)}{4\pi R_o}$$  \hspace{1cm} (A2.2)

Eq. (A2.1) essentially states that, in order to know the value of the sound pressure at the observation point $\overrightarrow{r}$, the value of the pressure and its gradient through any closed surface surrounding the observation point need to be known.
Appendix 2: Mathematical development of the diffraction/reflection model

Notice that $V_0g(\omega,R_o)$ in Eq. (A2.1) can be expressed as:

$$V_0g(\omega,R_o) = \cos \gamma_o \frac{\partial}{\partial R_o} g(\omega,R_o) = \cos \gamma_o \left( jk - \frac{1}{R_o} \right) g(\omega,R_o) \quad (A2.3)$$

where $\vec{n}_o$ is the unit vector normal to the surface (defined pointing outwards) and $\gamma_o = \text{angle}(\vec{n}_o,R_o)$, with $R_o = r_o - \vec{r}$.

A2.2 Diffracted pressure at any point behind the aperture

Eq. (A2.1) is the main result used to derive Kirchoff's theory of diffraction and its approximation to obtain the diffracted pressure behind apertures (Braddock, 1965; Ditchburn, 1976; Hecht, 1990). For an incident plane wave of the form shown in Eq. (5.3), the value of $V_0p(r_o)$ can be expressed as:

$$V_0p(r_o) = jk \cos(\vec{n}_o, \vec{k}) p(r_o) \quad (A2.4)$$

The diffracted pressure at any point behind the aperture [i.e., Eq. (5.6)] is obtained by substituting Eqs. (A2.2), (A2.3) and (A2.4) and the value of the incident plane wave at each point on the aperture into Eq. (A2.1). The expression for the diffraction obliquity factor $K$ [Eq. (5.8)] emerges naturally from the calculations. The observation points, in our particular case, are all the points on the reflecting surface (i.e., $\vec{r} = r_q$, $\forall r_q \in S_R$). For the case of diffraction by an aperture, the integration surface, $S$, becomes the surface of the aperture, $S_A$, and the zero subscript, which refers to a general arbitrary surface, must be substituted by the subscript $l$, which specifically refers to $S_A$. Therefore, the vector $\vec{R}_o = \vec{r}_o - \vec{r}$ reduces to $r_{ql}$, the angle $\gamma_o$ becomes $\theta_{ql}$ and $\cos(\vec{n}_o, \vec{k}) = \cos \psi_l$ (see Fig. 5.1a).

---

1 Kirchoff's analysis applies to an integral taken over the whole of a closed surface which surrounds the point at which we are interested in finding the diffracted wave. When an obstacle is inserted, (for instance, an infinitely large opaque screen in which there is an aperture) the pressure field at the point of interest can be calculated by integrating over the unobstructed part of the wavefront, i.e., over the surface of the aperture. This approximation is normally known as St. Venant's hypothesis (Ditchburn, 1976, p.174).
A2.3 Reflected pressure from the reflecting wall

The general result expressed in Eq. (A2.1) has been also applied to obtain the reflected pressure from the concha reflecting surface [i.e., to obtain Eq. (5.10)]. However, in this case, the integration must be done over the reflecting surface, \( S_R \), and boundary conditions on the surface must be applied. In addition, the subscript zero must be substituted by \( q \), which specifically refers to \( S_R' \). The observation point for the reflection process is the meatus entrance, which, for convenience, has been defined at the origin of the system of co-ordinates (i.e., \( \vec{r} = 0 \)). Therefore, \( \vec{R}_o = \vec{r}_o \) which has been named \( \vec{r}_q \) in our case (see Fig. 5.1a).

If the reflecting surface is such that total reflection can be assumed (i.e., it is perfectly rigid), the boundary condition is that the normal component of the particle velocity must be zero at the reflecting surface (Morse and Ingard, 1968, p. 366). This condition is held if, and only if:

\[
\nabla_q p(\vec{r}_q) = 0 \quad (A2.5)
\]

Therefore, with the above considerations Eq. (A2.1) reduces to:

\[
\mathbf{p}_R = -\iiint_{S_R} [p(\vec{r}_q) \nabla_q s(\omega, r_q)] ds_q \quad (A2.6)
\]

The previous expression requires the value of the total pressure field on the reflecting surface (i.e., \( p(\vec{r}_q) \)) to be known. In our case, \( p(\vec{r}_q) \) has been approximated as the diffracted pressure on the reflecting surface, \( p_d(\vec{r}_q) \) (approximation (iv)). Under this assumption, the reflected pressure at the meatus entrance can be obtained by substituting Eq. (A2.3) into (A2.6). The expression for the reflection obliquity factor for the case of total reflection [Eq. (5.11)], \( R_{(r_q, r_q)} \), is obtained trivially from the calculations as:
Appendix 2: Mathematical development of the diffraction/reflection model

\[
R(q, \gamma_q) = \frac{\nabla_{\alpha} g(\omega, R_q)}{4\pi \cdot g(\omega, R_q)} = \frac{\cos \gamma_q}{4\pi} \left( \frac{1}{r_q} - jk \right) \tag{A2.7}
\]

If the reflecting surface is not perfectly rigid but is still passive, the boundary condition is no longer the one given by Eq. (A2.5) but (Morse and Ingard, 1968):

\[
\nabla_q p(r_q) = j\omega \rho_q \frac{p(r_q)}{Z_N(r_q, \omega)} \tag{A2.8}
\]

where \( Z_N(r_q) \) is the normal acoustic impedance of the surface and \( \rho_q \) is the density of the reflecting material. \( Z_N(r_q) \) is defined as the ratio between the pressure and the normal component of the particle velocity at the reflecting surface, i.e.:

\[
Z_N(r_q) = \frac{p(r_q)}{u_N(r_q)} \tag{A2.9}
\]

For real materials, \( Z_N(r_q) \) varies with the frequency of the incident sound wave.

Substituting Eq. (A2.8) into Eq. (A2.1), considering approximation (iv) and operating we obtain an approximate value for the total reflected pressure at the origin:

\[
P_R = \iint_{S_R} \left[ \frac{p(r_q)}{Z_N(r_q)} \left( \frac{j\omega \rho_q}{Z_N(r_q)} - j\omega \gamma_q \left( \frac{1}{r_q} - jk \right) \right) \right] d\sigma_q \tag{A2.10}
\]

In this case, the reflection obliquity factor is:

\[
R(q, \gamma_q, Z_N) = \frac{j\omega \rho_q}{4\pi Z_N(r_q)} - \frac{\cos \gamma_q}{4\pi} \left( \frac{1}{r_q} - jk \right) \tag{A2.11}
\]

Now \( R(q, \gamma_q, Z_N) \) not only depends on the geometry of the problem but also on the acoustic properties of the reflecting surface. Note that, when \( Z_N = \infty \) (i.e., total reflection), the value of \( R(q, \gamma_q, Z_N) \) reduces to \( R(q, \gamma_q) \), as it should happen.
APPENDIX 3

Computer code for evaluating the diffraction/reflection model

The computer code employed for evaluating the diffraction/reflection model is given below. The code is written in ANSI C programming language.

Two programs are included: MetalC7b.c, and 3DConcha10.c. The first program, MetalC7b.c, evaluates the diffraction/reflection model for the metallic spiral shown in Fig. 5.1a. The second program evaluates the diffraction/reflection model for the approximated concha shape shown in Fig. 5.3. Both programs use some of functions provided with LUTEar (LUTEar, 1993). Note that these functions have not included in the code listings shown below, as they can be simply obtained by downloading LUTEar [the internet address is specified in the reference LUTEar (1993)]. The parameters used to obtained the results presented in Chapter 5 are given at the end of each program.

Geometrical considerations

For the diffraction/reflection model to be evaluated the surface of the diffracting aperture and the reflecting surface must be described. Both surfaces are generated within the code, from the YZ contours of the metallic spiral and the realistic concha shape. Each surface is described as a collection of surface differentials (ds). In the code, each surface differential is described by a vector position, the area of the surface element, and the normal vector to that surface element. Figs. A3.1 and A3.2 illustrate the geometrical variables defined in the code for the metallic spiral and for the realistic concha. The variable names used in the code are preserved in the figure.

---

Figure A3.1. An illustration of the geometrical variables employed in the C program MetalC7b.c.
Appendix 3: Computer code

Figure A3.2. An illustration of the geometrical variables employed in the C program 3DConcha10.c.
Appendix 3: Computer code

 /************************************************************************
 */
 */ MetalC7b.c 
 */
 */ This program finds the transfer function of the Metal Spiral.
 */ It is a working version which uses the right obliquity factor.
 */ It uses a collection of vectors defining the diffracting aperture
 */ and a collection of vectors defining the reflecting surface.
 */ The program runs the 3D version or the 2D approximation depending
 */ on the value of the parameters (length of the
 */ diffracting/reflecting spiral).
 */
 */ Some of the geometrical variables defined in the code are
 */ illustrated in Fig. A3.1.
 */
 */ A parameter file is given at the end of the program.
 */
 */ This code only runs one source position at a time.
 */
 */ © Enrique A. Lopez-Poveda, 1996.
 */
************************************************************************/

 #include <stdio.h>
 #include <stdlib.h>
 #include <math.h>
 #include "GeLUTEar.h" /* LUTEar library */
 #include "XYData.h" /* LUTEar library */
 #include "UtVectorM.h" /* LUTEar library */
 #include "UtCmplxM.h" /* LUTEar library */
 #include 'FiParFile.h' /* LUTEar library */

 /*********************************************************************/
 /*********************************************************************/
 /**************************************************************************/
 /**************************************************************************/

 #define PARAMETERS_FILE "MetalC7b.par" /* Name of paramters file. */
 #define MAX_NUMBER_OF_DS 20 /* max No of ds for diffraction */
 #define MIN_NUMBER_OF_DS 1 /* min No of ds for diffraction */
 #define MAX_NUMBER_OF_DSR 20 /* max No of ds for reflection */
 #define MIN_NUMBER_OF_DSR 1 /* minimum */
 #define MAX_NUMBER_OF_BANDS 20 /* max No of bands */
 #define MIN_NUMBER_OF_BANDS 10 /* min No of bands */
 #define MAX_NUMBER_OF_REFLECTING_DS 2000
 #define SPEED 33210.0 /* speed of sound */

 /*********************************************************************/
 /*********************************************************************/
 /**************************************************************************/
 /**************************************************************************/

typedef struct {
    Vector vR; /* vector position of a reflecting differential */
    Vector nR; /* unit-vector normal to a reflecting differential */
    double dRArea; /* area of a reflecting differential */
} dRef, *dRefPtr;

typedef struct {
    Vector vD; /* vector position of a diffracting differential */
    Vector nD; /* unit-vector normal to a diffracting differential */
} dDiff, *dDiffPtr;

281
double dDArea; /* area a diffracting differential */

/* Global variables */

char outputFile[MAXLINE], inputShapeFile[MAXLINE];
int numBands, numdS, numdSr;
double minFreq, maxFreq, freqStep, reflectCoeff_Re, reflectCoeff_Im,
degAzim, degElev, apertureLength;

/* Functions and subroutines */

/* ReadParsFromfile */

This routine reads a specified number of parameters from a file.
It expects there to be one parameter per line.
This is LUTEar function.

void ReadParsFromfile(char *fileName)
{
    char line[MAXLINE];
    FILE *fp;

    if ((fp = fopen(fileName, "r"); == NULL) {
        NotifyError("ReadTestPars: Cannot open data file '", fileName);
        exit(1);
    }

    printf("Reading parameters from file: %s\n", fileName);
    Init_ParFile();
    GetPars_ParFile(fp, "%s", outputFile);
    GetPars_ParFile(fp, "%s", inputShapeFile);
    GetPars_ParFile(fp, "%lf", &apertureLength);
    GetPars_ParFile(fp, "%d", &numBands);
    GetPars_ParFile(fp, "%d", &numdS);
    GetPars_ParFile(fp, "%d", &numdSr);
    GetPars_ParFile(fp, "%lf", &reflectCoeff_Re);
    GetPars_ParFile(fp, "%lf", &reflectCoeff_Im);
    GetPars_ParFile(fp, "%lf", &minFreq);
    GetPars_ParFile(fp, "%lf", &maxFreq);
    GetPars_ParFile(fp, "%lf", &freqStep);
    GetPars_ParFile(fp, "%lf", &degAzim);
    GetPars_ParFile(fp, "%lf", &degElev);
    fclose(fp);
    Free_ParFile();
}
/********************************************************* CheckPars *********************************************************/ /* * This program checks the read parameters. * This is a LUTEar function * */

void CheckPars ()
{
    if ((numBands < MIN_NUMBER_OF_BANDS) ||
         (numBands > MAX_NUMBER_OF_BANDS))
    {
        printf ("\nnumBands must be greater than %d and fewer than "/
                " %d!\n", MIN_NUMBER_OF_BANDS, MAX_NUMBER_OF_BANDS);
        exit (1);
    }
    if ((numdS < MIN_NUMBER_OF_DS) || (numdS > MAX_NUMBER_OF_DS))
    {
        printf ("\numndS must be at greater than %d and fewer than "/
                " %d!\n", MIN_NUMBER_OF_DS, MAX_NUMBER_OF_DS);
        exit (1);
    }
    if ((numdSr < MIN_NUMBER_OF_DSR) || (numdSr > MAX_NUMBER_OF_DSR))
    {
        printf ("\numndSr must be at greater than %d and fewer than "/
                " %d!\n", MIN_NUMBER_OF_DS, MAX_NUMBER_OF_DS);
        exit (1);
    }
    if (apertureLength <= 0)
    {
        printf ("\nIllegal value of apertureLength!\n");
        exit (1);
    }
    if (maxFreq <= minFreq)
    {
        printf ("\n Illegal value of maxFreq!\n");
        exit (1);
    }
    if (minFreq < 1.0)
    {
        printf ("\nIllegal value of minFreq!\n");
        exit (1);
    }
    if (freqStep <= 0.0)
    {
        printf ("\nIllegal value of freqStep!\n");
        exit (1);
    }
    if ((degElev < -50.0) || (degElev > 90.0))
    {
        printf ("\nIllegal value of elevation!\n");
        exit (1);
    }
    if ((degAzim < 0.0) || (degAzim > 90.0))
    {
        printf ("\nIllegal value of azimuth!\n");
        exit (1);
    }
    if ((degAzim > 0.0))
    {
        printf ("\nAzimuth angle different from zero!\n");
        printf ("\tValue not appropriate for the metal concha!\n");
        exit (1);
    }
}
Appendix 3: Computer code

/***************************************************************************/
/*
* This routine prints the read parameters.
*/
/***************************************************************************/

void PrintPars()
{
    printf( "\tShape read from: %s\n", inputShapeFile);
    printf( "\tMetallic surface length: %.llf\n", apertureLength);
    printf( "\tNumber of diffracting bands: %d\n", numBands);
    printf( "\tNumber of diffracting differentials in each band: %d\n", numdS);
    printf( "\tNumber of reflectin differentials in each reflecting band: %d\n", numdSr);
    printf( "\tReflection Coefficient: %.llf + j %.llf\n", reflectCoeff_Re, reflectCoeff_Im);
    printf( "\tMaximum frequency: %.llf\n", maxFreq);
    printf( "\tMinimum frequency: %.1lf\n", minFreq);
    printf( "\tFrequency step: %.1lf\n", freqStep);
    printf( "\tAzimuth: %.1lf\n", degAzim);
    printf( "\tElevation: %.llf\n", degElev);
}

/***************************************************************************/
/* This program prints the read parameters to file. */
/***************************************************************************/

void PrintParsToFile(FILE *fp, double degElev, double degAzim)
{
    fprintf(fp, "Metal Spiral Model\n");
    fprintf(fp, "Code used: MetaICI.c\n");
    fprintf(fp, "Obliq. factor: NEW #405\n");
    fprintf(fp, "Shape read from: %s\n", inputShapeFile);
    fprintf(fp, "Maximum frequency: %.llf\n", maxFreq);
    fprintf(fp, "Minimum frequency: %.llf\n", minFreq);
    fprintf(fp, "Frequency step: %.llf\n", freqStep);
    fprintf(fp, "Number of diffracting bands: %d\n", numBands);
    fprintf(fp, "Number of aperture differentials per band: %d\n", numdS);
    fprintf(fp, "Number of reflecting differentials per band: %d\n", numdSr);
    fprintf(fp, "Reflection Coefficient: %.1lf + j %.1lf\n", reflectCoeff_Re, reflectCoeff_Im);
    fprintf(fp, "Elevation angle: %.1lf\n", degElev);
    fprintf(fp, "Azimuth angle: %.1lf\n", degAzim);
}

/***************************************************************************/
/* This function calculates the position of all the diffracting differentials in the aperture. */
/***************************************************************************/
* The input to the function is the contour of the metallic spiral. *

void CalcDiffractingDifferentials (dDifPtr diffSface, XYDataPtr shape, 
int numdS, int numBands) {
    int i, j, k;
    double apertureWidth, dX, dZ;

    apertureWidth = fabs(shape->y[0]) + 
                    fabs(shape->y[shape->numPoints-1]);
    dZ = apertureWidth / numBands;
    dX = apertureLength / numdS;
    i = 0;
    for ( j = 0; j < numdS; j++) {
        for (k = 0; k < numBands; k++) {
            diffSface[i].vD.x = j * dX;
            diffSface[i].vD.y = 0.0;
            diffSface[i].vD.z = shape->y[0] + k * dZ;
            diffSface[i].nD.x = 0.0;
            diffSface[i].nD.y = -1.0;
            diffSface[i].nD.z = 0.0;
            diffSface[i].dDArea = dZ * dX;
            i++
        } /* for k... */
    } /* for j... */
}

/*********************** CalcReflectingDifferentials *******************/

/*
* This function calculates the position of all the reflecting 
* differentials in the reflecting surface.
*/

void CalcReflectingDifferentials (dRefPtr refSface, XYDataPtr shape, 
int numdSr) {
    int i, j, k;
    double dX, modds;
    Vector ds;

    dX = apertureLength / numdSr;
    i = 0;
    for ( j = 0; j < numdSr; j++) {
        for (k = 0; k < shape->numPoints - 1; k++) {
            refSface[i].vR.x = j * dX;
            refSface[i].vR.y = shape->x[k];
            refSface[i].vR.z = shape->y[k];
            ds.x = 0.0;
            ds.y = shape->x[k+1] - shape->x[k];
            ds.z = shape->y[k+1] - shape->y[k];
            modds = Convert_VectorM (&ds);
            /* normal to reflecting sface, defined outwards */
            refSface[i].nR.x = 0.0;
            refSface[i].nR.y = ds.z / modds;
            refSface[i].nR.z = - ds.y / modds;
            refSface[i].dRArea = modds * dX;
            i++;
        } /* for k... */
    } /* for j... */
Appendix 3: Computer code

```cpp
void DiffractionBy_Jth_Differential(ComplexPtr diffPresJ, dDif positionVec, dRef targetVec, Vector iD, double kK) {
    double modrij, cos_r_n, cos_k_n, hh, delayj, rr, coeffj;
    Complex correc, expo;
    Vector rij;

    Subt_VectorM(&targetVec.vR, &positionVec.vD, &rij);
    modrij = Convert_VectorM(&rij);
    Power_VectorM(&rij, -1); /* to find out rji */
    cos_r_n = EscMult_VectorM(&positionVec.nD, &rij) / modrij;
    cos_k_n = EscMult_VectorM(&positionVec.nD, &iD);
    hh = 0.5 * (cos_r_n - cos_k_n);

    correc.re = hh * kK / (2 * PI);
    correc.im = 0.0;
    delayj = EscMult_VectorM(&iD, &positionVec.vD);
    rr = modrij + delayj;
    expo.re = positionVec.dDArea * cos(kK * rr - PI/2);
    expo.im = positionVec.dDArea * sin(kK * rr - PI/2);

    /* attenuation with distance for the diffracted wave is NOT
    * considered. If it was it should go right above in the expo as:
    * expo/modrij */
    Mult_CmplxM (&correc, &expo, diffPresJ);
}
```

```cpp
/**
 * The function calculates the diffracted pressure caused by a single
 * differential in the aperture (placed at the positionVec.vD) on a
 * reflecting differental placed at the targetVec.vD.
 */
void DiffractionBy_Jth_Differential(ComplexPtr diffPresJ, dDif positionVec, dRef targetVec, Vector iD, double kK) {
    double modrij, cos_r_n, cos_k_n, hh, delayj, rr, coeffj;
    Complex correc, expo;
    Vector rij;

    Subt_VectorM(&targetVec.vR, &positionVec.vD, &rij);
    modrij = Convert_VectorM(&rij);
    Power_VectorM(&rij, -1); /* to find out rji */
    cos_r_n = EscMult_VectorM(&positionVec.nD, &rij) / modrij;
    cos_k_n = EscMult_VectorM(&positionVec.nD, &iD);
    hh = 0.5 * (cos_r_n - cos_k_n);

    correc.re = hh * kK / (2 * PI);
    correc.im = 0.0;
    delayj = EscMult_VectorM(&iD, &positionVec.vD);
    rr = modrij + delayj;
    expo.re = positionVec.dDArea * cos(kK * rr - PI/2);
    expo.im = positionVec.dDArea * sin(kK * rr - PI/2);

    /* attenuation with distance for the diffracted wave is NOT
    * considered. If it was it should go right above in the expo as:
    * expo/modrij */
    Mult_CmplxM (&correc, &expo, diffPresJ);
}
```

```cpp
/**
 * This function calculates the pressure reflected at the origin from a
 * single surface differential placed at targetVec.vR the reflecting
 * surface.
 * It needs the total incident diffracted pressure on that element as the
 * input.
 */
```
Appendix 3: Computer code

```c
void
ReflectionBy_Ith_Differential (ComplexPtr refPresI, dRef targetVec,
    ComplexPtr totalDiffPres, double kK) {
    double modri, cos_r_n;
    Complex expo2, correc;

    modri = Convert_VectorM(&targetVec.vR);
    /* attenuation with distance for the reflected wave is considered
     * (see the expo2) */
    expo2.re = targetVec.dRArea * cos (kK * modri) / modri;
    expo2.im = targetVec.dRArea * sin (kK * modri) / modri;
    cos_r_n = EscMult_VectorM(&targetVec.nR, &targetVec.vR) / modri;

    /* to change the condition under which the diff/ref model is
     * evaluated for the reflection obliquity factor, only the values of
     * correc.re and correc.im need to be changed.
     * To consider attenuation of the reflected wave with distance can
     * be specified in expo2 (see above), for instance. */
    correc.re = cos_r_n / (4 * PI * modri);
    correc.im = - cos_r_n * kK / (4 * PI);
    Mult_CmplxM (&expo2, &correc, &expo2);
    Mult_CmplxM (&expo2, totalDiffPres, refPresI);
}
```

```c
void
main () {
    XYDataPtr data = NULL;
    int i, j, totaldS, totaldSr;
    double elev, azim, freq, kK, wavelength;
    FILE *fp;
    dDiffPtr diffractingSface;
    dRefPtr reflectingSface;
    Complex diffPresj, totalDifPres, refPresi, totalRefPres,
        refCoeff, tranFunct, modulus;
    Vector id;

    /* Reading, checking and printing the parameters */
    ReadParsFromFile (PARAMETERS_FILE);
    CheckPars();
    PrintPars();

    /* Memory allocation for the differentials of diffracting surface */
    totaldS = numdS * numBands;
    if ((diffractingSface = (dDiffPtr) malloc(sizeof(dDiff) * totaldS))
        == NULL) {
        fprintf(stderr, "Out of Memory");
    
    } /* Main process */
}
```

exit(1);
}

/* Reading the surface contour ---------------------------------------------*/
if(!Read_XYData(inputShapeFile,&data))
{
    fprintf(stderr, "Could not read file %s.\n",inputShapeFile);
    exit(1);
}
printf ("\tTotal number of diffracting differentials %d: \n",totaldS);
totaldSr = numdSr * (data->numPoints - 1);
printf ("\tTotal number of reflecting differentials %d: \n",totaldSr);

/* Memory allocation for the differentials of reflecting surface */
if ((reflectingSface = (dRefPtr) malloc(sizeof(dRef) * totaldSr))
    == NULL) {
    fprintf(stderr, "Out of Memory\n");
    exit(1);
}
if ((fp = fopen(outputFile, "w+")) == NULL) {
    printf("Can't open output file %s.\n", outputFile);
    exit(1);
}
PrintParsToFile (fp, degElev, degAzim);

/* Calculating the diffracting and reflecting differentials that * define the aperture and the reflecting surface of the system */
CalcDiffractingDifferentials (diffractingSface, data, numdS, numBands);
CalcReflectingDifferentials (reflectingSface, data, numdSr);

/* Initialisation of variables --------------------------------------------*/
elev = degElev * (PI / 180); /* converting degElev to radians */
azim = degAzim * (PI / 180); /* converting degAzim to radians */

/* Calculating the vector that defines the incident direction of * the sound from the elev and the azim angles. * iD is the unit incident direction vector --------------------------------/
iD.x = -cos(elev) * sin(azim);
iD.y = cos(elev) * cos(azim);
iD.z = -sin(elev);
refCoeff.re = reflectCoeff_Re;
refCoeff.im = reflectCoeff_Im;

/* Evaluating the diff/ref model for each frequency ---------------------*/
for (freq = minFreq; freq <= maxFreq + freqStep; freq += freqStep ){
    wavelength = SPEED / freq;
    KK = (2.0 * PI) / wavelength; /* wavenumber */
    for (i = 0, totalRefPres.re = 0.0, totalRefPres.im = 0.0;
        i < totaldSr; i++){
        /* Calculating the total diffracted pressure over the i-th * reflecting element */
        for (j = 0, totalDifPres.re = 0.0, totalDifPres.im = 0.0;
            j < totaldS; j++) {
Appendix 3: Computer code

```c
DiffractionBy_Jth_Differential(&diffPresj,
   diffractingSface[j], reflectingSface[i], iD, kK);
Add_CmplxM (&totalDifPres, &diffPresj, &totalDifPres);
/*
 * totalDifPres gives the total diffracted pressure
 * on a reflecting differential
 */
} /* for j...*/

/* Calculating the reflected pressure from the i-th
 * element over the origin (microphone) */
ReflectionBy_Ith_Differential (&refPresi,
   reflectingSface[i], &totalDifPres, kK);

/* Adding the reflected pressure from every reflecting
 * differential */
Add_CmplxM (&totalRefPres, &refPresi, &totalRefPres);
} /* for i...*/

Mult_CmplxM(&totalRefPres, &refCoeff, &totalRefPres);

/* Calculating the transfer function (see Eq. 5.13) */
totalRefPres.re += 1.0;
totalRefPres.im = totalRefPres.im;

/* Calculating the modulus of the transfer function */
Convert_CmplxM(&totalRefPres, &tranFunct);

/* printing results out */
fprintf(fp, "%f	%f
", freq, 20 * log10(tranFunct.re));
} /* for freq... */
printf("\nOutput sent to %s file!\n", outputFile);
fclose(fp);
}

/**************************** End of MetalC7b.c *************************/
Parameter file: MetalC7b.par

This is the parameter file employed for obtaining the results shown in Fig. 5.2 (i.e., results for the metal spiral)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>metal.out</td>
<td>Output file name.</td>
</tr>
<tr>
<td>input2SizeSpiral180_data.dat</td>
<td>Input shape file name.</td>
</tr>
<tr>
<td>6.0</td>
<td>Aperture length (cm).</td>
</tr>
<tr>
<td>20</td>
<td>Number of diffracting and reflecting bands.</td>
</tr>
<tr>
<td>5</td>
<td>Number of diffracting differentials per band.</td>
</tr>
<tr>
<td>5</td>
<td>Number of reflecting differentials per reflecting band.</td>
</tr>
<tr>
<td>1.0</td>
<td>Real part of the reflection coef.</td>
</tr>
<tr>
<td>0.0</td>
<td>Imaginary part of the reflection coef.</td>
</tr>
<tr>
<td>1</td>
<td>Minimum frequency (Hz).</td>
</tr>
<tr>
<td>17000</td>
<td>Maximum frequency (Hz).</td>
</tr>
<tr>
<td>250</td>
<td>Frequency Step (Hz).</td>
</tr>
<tr>
<td>0</td>
<td>Azimuth angle (degrees, between 0 and 90)</td>
</tr>
<tr>
<td>-45.0</td>
<td>Elevation angle (degrees, between - 40 and 90)</td>
</tr>
</tbody>
</table>

The file 'input2SizeSpiral180_data.dat' contained 180 data-pairs that define the YZ contour (Fig. 5.1) of the metallic spiral (see also Fig. A3.1).
The program evaluates the diffraction/reflection model for the approximated concha shape given in Fig. 5.3. (See Chapter 5)

Some of variables used in the code are specified in Fig. A3.2.

This program evaluates the diffraction/reflection model for as many azimuthal and elevation angles as specified in the parameter file.

© Enrique A. Lopez-Poveda, 1996.

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include "GeLUTEar.h" /* LUTEar library */
#include "XYData.h" /* LUTEar library */
#include "UtVectorM.h" /* LUTEar library */
#include "UtCmplxM.h" /* LUTEar library */
#include "FiParFile.h" /* LUTEar library */

#define PARAMETERS_FILE "3DConcha10.par" /* Name of params. file */
#define MAX_NUMBER_OF_BANDS 60 /* max No of diffracting bands */
#define MIN_NUMBER_OF_BANDS 20 /* min No of diffracting bands */
#define MIN_NUMBER_OF_DS 20 /* min No. of ds within a band */
#define SPEED 33210.0 /* speed of sound */

typedef struct {
    double y; /* see Fig. A3.2 */
    double xrn; /* see Fig. A3.2 */
    double xM; /* see Fig. A3.2 */
} IntegrationBand;

typedef struct {
    double ymin; /* see Fig. A3.2 */
    double ymax; /* see Fig. A3.2 */
    double area; /* see Fig. A3.2 */
    double widthOfBands; /* see Fig. A3.2 */
} Concha, *ConchaPtr;

...
Appendix 3: Computer code

char outputFile[MAXLINE], inputShapeFile[MAXLINE];
int numBands, numDS;
double minFreq, maxFreq, freqStep, conchaTilt, conchaDepth;
double reflectCoeff_Re, reflectCoeff_Im, minElevation, maxElevation;
double elevStep, minAzimuth, maxAzimuth, azimStep;

/******************************************************************************
/******************************************************************************
/**************************** Functions and subroutines *************************
/******************************************************************************
/******************************************************************************

/**************************** ReadParsFromFile *******************************/

void ReadParsFromFile(char *fileName)
{
    char line[MAXLINE];
    FILE *fp;

    if ((fp = fopen(fileName, "r")) == NULL) {
        NotifyError("ReadTestPars: Cannot open data file '%s'.
        file\n        exit(1);
    }
    printf("Reading parameters from file: %s\n", fileName);
    Init_ParFile();
    GetPars_ParFile(fp, "%s", outputFile);
    GetPars_ParFile(fp, "%s", inputShapeFile);
    GetPars_ParFile(fp, "%d", &numBands);
    GetPars_ParFile(fp, "%d", &numDS);
    GetPars_ParFile(fp, "%lf", &conchaTilt);
    GetPars_ParFile(fp, "%lf", &conchaDepth);
    GetPars_ParFile(fp, "%lf", &reflectCoeff_Re);
    GetPars_ParFile(fp, "%lf", &reflectCoeff_Im);
    GetPars_ParFile(fp, "%lf", &minFreq);
    GetPars_ParFile(fp, "%lf", &maxFreq);
    GetPars_ParFile(fp, "%lf", &freqStep);
    GetPars_ParFile(fp, "%lf", &minElevation);
    GetPars_ParFile(fp, "%lf", &maxElevation);
    GetPars_ParFile(fp, "%lf", &elevStep);
    GetPars_ParFile(fp, "%lf", &minAzimuth);
    GetPars_ParFile(fp, "%lf", &maxAzimuth);
    getclose(fp);
    Free_ParFile();
}

/**************************** CheckPars *************************************/

/* This routine checks that the read parameters have adequate values.
*/

292
void CheckPars ()
{
    if ((numBands < MIN_NUMBER_OF_BANDS) ||
        (numBands > MAX_NUMBER_OF_BANDS)) {
        printf("\nnumBands must be greater than %d and fewer than %d!\n", 
               MIN_NUMBER_OF_BANDS, MAX_NUMBER_OF_BANDS);
        exit (1);
    }
    if (numDS < MIN_NUMBER_OF_DS) {
        printf("\numDS must be at greater than %d!\n", MIN_NUMBER_OF_DS);
        exit (1);
    }
    if ((conchaTilt < -0.0) || (conchaTilt > 90.0)) {
        printf("\nIllegal value of conchaTilt!\n");
        exit (1);
    }
    if (conchaDepth <= 0.0) {
        printf("\nIllegal value of conchaDepth!\n");
        exit (1);
    }
    if (maxFreq <= minFreq) {
        printf("\nIllegal value of maxFreq!\n");
        exit (1);
    }
    if (minFreq < 1.0) {
        printf("\nIllegal value of minFreq!\n");
        exit (1);
    }
    if (freqStep <= 0.0) {
        printf("\nIllegal value of freqStep!\n");
        exit (1);
    }
    if ((minElevation < -50.0) || (maxElevation > 90.0)) {
        printf("\nIllegal value of either minElevation or maxElevation\n");
        exit (1);
    }
    if ((minAzimuth < 0.0) || (maxAzimuth > 90.0)) {
        printf("\nIllegal value of either minAzimuth or maxAzimuth\n");
        exit (1);
    }
    if ((elevStep <= 0.0) || (azimStep <= 0.0)) {
        printf("\nIllegal value of either elevStep or azimStep\n");
        exit (1);
    }
}

/************************ PrintPars ****************************/
/*
 * This routine prints the read parameters.
 */

void PrintPars ()
{
    printf("\tShape read from: %s\n", inputShapeFile);
    printf("\tNumber of diffracting bands: %d\n", numBands);
    printf("\tNumber of diffracting differentials in each band: %d\n", 

Appendix 3: Computer code

```c
numdS);
printf ("\tConcha tilt: %.2lf\n", conchaTilt);
printf ("\tConcha depth: %.2lf\n", conchaDepth);
printf ("\tReflection Coefficient: %.1lf + j %.1lf\n",
    reflectCoeff_Re, reflectCoeff_Im);
printf ("\tTotal frequency: %.1lf\n", maxFreq);
printf ("\tMinimum frequency: %.1lf\n", minFreq);
printf ("\tFrequency step: %.1lf\n", freqStep);
printf ("\tMaximum elevation: %.1lf\n", maxElevation);
printf ("\tMinimum elevation: %.1lf\n", minElevation);
printf ("\tElevation step: %.1lf\n", elevStep);
printf ("\tMaximum azimuth: %.1lf\n", maxAzimuth);
printf ("\tMinimum azimuth: %.1lf\n", minAzimuth);
printf ("\tAzimuth step: %.1lf\n", azimStep);
}

/*********************************************************
** PrintParsToFile **********************************************************/
*/
void
PrintParsToFile (FILE *fp, double degElev, double degAzim)
{
    fprintf (fp,"3D CONCHA MODEL\n");
    fprintf (fp,"Code used: 3DConchal.c\n");
    fprintf (fp,"Obliq. factor: NEW #405\n");
    fprintf (fp,"Shape read from: %s\n", inputShapeFile);
    fprintf (fp,"Maximum frequency: %.1lf\n", maxFreq);
    fprintf (fp,"Minimum frequency: %.1lf\n", minFreq);
    fprintf (fp,"Frequency step: %.1lf\n", freqStep);
    fprintf (fp,"Number of diffracting bands: %d\n", numBands);
    fprintf (fp,"Number of aperture differentials per band: %d\n", numdS);
    fprintf (fp,"Concha tilt angle: %.2lf\n", conchaTilt);
    fprintf (fp,"Concha depth: %.2lf\n", conchaDepth);
    fprintf (fp,"Reflection Coefficient: %.1lf + j %.1lf\n",
        reflectCoeff_Re, reflectCoeff_Im);
    fprintf (fp,"Elevation angle: %.1lf\n", degElev);
    fprintf (fp,"Azimuth angle: %.1lf\n", degAzim);
    fprintf (fp,"freq (Hz.)\ntpressure\n");
}

/*********************************************************
** StraightLine **********************************************************/
*/
double
StraightLine(double x1, double y1, double x2, double y2, double y)
{
    double x;
    x = ((x2 - x1)/(y2 - y1)) * (y - y1) + x1;
    return (x);
}
```

Appendix 3: Computer code

1************************ ClosedIntegration_Coeff ***********************/

/*
* This function returns the value of the integration coefficients for an
* closed integral. The value of the coefficients
* were taken from Numerical Recipes in C (Press et al., 1992).
*/

double ClosedIntegration_Coeff(int i_th, int max_i)
{
    double coeff;
    if ((i_th == 0) || (i_th == max_i - 1)) coeff = 3.0 / 8.0;
    else if ((i_th == 1) || (i_th == max_i - 2)) coeff = 7.0 / 6.0;
    else if ((i_th == 2) || (i_th == max_i - 3)) coeff = 23.0 / 24.0;
    else coeff = 1.0;
    return (coeff);
}

/************************ OpenIntegration_Coeff ************************/

/*
* This function returns the value of the integration coefficients for an
* open integral. The value of the coefficients were taken
*/

double OpenIntegration_Coeff(int i_th, int max_i)
{
    double coeff;
    if ((i_th == 1) || (i_th == max_i - 2)) coeff = 55.0 / 24.0;
    else if ((i_th == 1) || (i_th == max_i - 3)) coeff = -1.0 / 6.0;
    else if ((i_th == 2) || (i_th == max_i - 4)) coeff = 11.0 / 8.0;
    else coeff = 1.0;
    return (coeff);
}

/************************ DiffractionByBand ****************************/

/*
* This function finds the diffracted pressure caused by a
* diffracting band, k, at a point on the concha contour, rq, for a
* given elevation, elev.
* The function returns the complex ptr Ql which is the total pressure
* diffracted by the k band of the diffracting aperture.
* kK : wavevector = 2PI/wavelength.
*/

void DiffractionByBand (ComplexPtr Ql, int k, Vector v, double elev,
                        double az, double kK, Concha *concha)
{
    int j;
    double bandTopLimitIn, bandLowLimitIn, temp, cos_r_n, hhQS,
            coeff, modsq, rr, dSa, delta, delayS, tilt, cos_k_n, modrqi;
    Vector rb, rs, D_rsq, rd, iD, D_rs, n_dSa;

    295
Complex  correc, expo;

tilt = conchaTilt * (PI / 180.0); /* convert tilt to radians */

if (concha->integrationBand[k].xM <= concha->integrationBand[k+1].xM){
    bandTopLimitIn = concha->integrationBand[k].xM;
} else if (concha->integrationBand[k].xM >
    concha->integrationBand[k+1].xM){
    bandTopLimitIn = concha->integrationBand[k+1].xM;
}
if (concha->integrationBand[k].xM <= concha->integrationBand[k+1].xM){
    bandLowLimitIn = concha->integrationBand[k+1].xM;
} else if (concha->integrationBand[k].xM >
    concha->integrationBand[k+1].xM) {
    bandLowLimitIn = concha->integrationBand[k].xM;
}
/* rb is defined in Fig. A3.2 */
rb.x = 0.0;
rb.y = bandLowLimitIn;
rb.z = concha->integrationBand[k].y;
delta = (bandTopLimitIn - bandLowLimitIn) / numdS;
/* rd is defined in Fig. A3.2 */
rd.x = 0.0;
rd.y = delta; /* see Fig. A3.2 */
rd.z = 0.0;
/* iD vector that defines the incident direction of the sound */
iD.x = - cos(elev) * sin(az);
iD.y = cos(elev) * cos(az);
iD.z = - sin(elev);
/* n_dSa is normal vector to the diffracting surface pointing out */
n_dSa.x = cos(tilt); /* n_dSa = normal vector to Sa outwards */
n_dSa.y = -sin(tilt);
n_dSa.z = 0.0;
/* dSa = area of the differential dSa of diffracting aperture */
dSa = (delta / cos(tilt)) * concha->widthOfBands;
for (j = 0, Ql->re = 0.0, Ql->im = 0.0; j < numdS; j++) {
    rd.x = 0.0;
    rd.y = delta;
    rd.z = 0.0;
    Power_VectorM(&rd, j);
    Add_VectorM(&rd, &rb, &rs); /* see Fig. A3.2 */
/* D_rs is the 3D position vector or a diffracting surface
   * differential */
    D_rs.y = rs.y;
    D_rs.z = rs.z;
    D_rs.x = rs.y * sin(tilt) / cos(tilt);
    Subt_Vectorm(&v, &D_rs, &D_rsq);
modsq = Convert_VectorM(&D_rsq);
Power_VectorM(&D_rsq, -1.0); /* to find -&D_rsq, ie., &D_rqs */
temp = EscMult_VectorM(&n_dSa, &D_rsq);
cos_r_n = temp / modsq;
cos_k_n = EscMult_VectorM(&n_dSa, &iD);

hhQS = 0.5 * (cos_r_n - cos_k_n);
correct.re = hhQS * kK / (2 * PI); /* kK/(2*PI) = wavelength */
correct.im = 0.0;
delayS = EscMult_VectorM(&iD, &D_rsq);

rr = modsq + delayS;

/*
* attenuation with distance not considered. Hence it is not
* dividing the variable "expo".
*/
expo.re = dSa * ClosedIntegration_Coeff (j, numdS)
* cos(kK * rr - PI / 2.0);
expo.im = dSa * ClosedIntegration_Coeff (j, numdS)
* sin(kK * rr - PI / 2.0);

Mult_CmplxM (&correct, &expo, &expo);
Add_CmplxM (Ql, &expo, Ql);

/* Ql gives us the total diffracted pressure by each band
* diffracting band.*/

void ReflectionBySurface (ComplexPtr refP, Vector rq, Vector rq_1,
ComplexPtr diffP, double kK, double concha_xMax) {

tilt = conchaTilt * (PI / 180.0); /* converts the tilt to radians */
approxDepth = concha_xMax * sin (tilt) / cos (tilt);
depthFactor = conchaDepth / approxDepth;
dX = rq.y * sin (tilt) / cos (tilt);
Subt_VectorM(&rq_1, &rq, &rqd);
modrq = Convert_VectorM(&rq);
modrqD = Convert_VectorM(&rqd);
dSr = modrqD * dX * depthFactor;

Appendix 3: Computer code
Appendix 3: Computer code

/* nSq = normal unity to the reflecting surface, pointing OUTWARD! */
#define nSq x 0.0;
#define nSq y = rqd z / modrqD;
#define nSq z = - rqd y / modrqD;
#define cos_r nSq = EscMult_VectorM (&rq, &nSq) / modrqD;

/* attenuation with distance not considered. Hence it is not dividing
* the variable expo2. */
#define expo2.re = cos (kK * modrq) * dSr;
#define expo2.im = sin (kK * modrq) * dSr;
#define correc.re = cos_r nSq;
#define correc.im = 0.0;
Mult_CmplxM (&expo2, &correc, &expo2);
Mult_CmplxM (&expo2, diffP, refP);

/***************************************************************************/
AnalysisConchaShape
/***************************************************************************/

/* The next function reads the points defining the concha contour and
* analyses the data to obtain the needed integration bands and the Ymin,
* Ymax and area values of the contour.
* Therefore, it returns the 'concha' structure. */

void AnalysisConchaShape (Concha *concha, XYDataPtr d) {

    int j, i, M, indexMax, indexMin, indexDiff, indexDiff2;
    double Xmax, Ymax, Xmin, Ymin, rI, rminDiff, minDiff2, yDiff, SL, 
inArea, outArea, topLimitIn, topLimitOut, lowLimitIn, lowLimitOut;

    M = numBands;

    if (M > d->numPoints) {
        printf ("\nNumber of diffracting bands exceeds the number\n           of reflecting points.\n'');
        printf ("EXECUTION ABORTED!!'');
        exit(1);
    }

    concha->ymin = 0.0; /* initialisation */
    concha->ymax = 0.0; /* initialisation */

    /* searching for ymax and ymin among the values in the contour */
    for (j=0; j < d->numPoints ; j++){
        if (concha->ymin <= d->y[j]) {
            concha->ymin = d->y[j];
            indexMin = j;
        }
        if (concha->ymax <= d->y[j]) {
            concha->ymax = d->y[j];
            indexMax = j;
        }
    }
    concha->ymax = d->y[indexMax];
    Xmax = d->x[indexMax];
Appendix 3: Computer code

concha->ymin = d->y[indexMin];
Xmin = d->x[indexMin];

printf ("concha->ymax = %f Xmax = %f \n", concha->ymax,
 d->x[indexMax]);
printf ("concha->ymin = %f Xmin = %f \n", concha->ymin,
 d->x[indexMin]);

/* defining the width of the bands in the concha aperture */
concha->widthOfBands = (concha->ymax - concha->ymin)/M;
printf ("Width of the Bands = %f\n", concha->widthOfBands);

/* searching for xm and xM for each band */
for (l = 0; l < M; l++) {
    minDiff = HUGE_VAL; /* conditions to initialize the comparisons */
    minDiff2 = HUGE_VAL;
    indexDiff = 0;
    indexDiff2 = 0;
    rl = concha->ymin + l*concha->widthOfBands;
    SL = StraightLine(Xmax, concha->yMax, Xmin, concha->yMin, rl);
    fprintf (fq,'%f	%f
', SL, rl);
    for (j=0; j < d->numPoints; j++) {
        if (d->x[j] >= SL) {
            if (fabs(d->y[j] - rl) < minDiff) {
                minDiff = fabs(d->y[j] - rl);
                indexDiff = j;
            }
        } else if (d->x[j] < SL) {
            if (fabs(d->y[j] - rl) < minDiff2)) {
                minDiff2 = fabs(d->y[j] - rl);
                indexDiff2 = j;
            }
        }
    }
    concha->integrationBand[l].xm = d->x[indexDiff2];
    concha->integrationBand[l].xM = d->x[indexDiff];
    concha->integrationBand[l].y = rl;
}

/* Next piece of code calculates the total are of concha aperture */
/* this area is not necessary for the implementation of the model */
/* The calculation is done for testing that everything is OK */

inArea = 0.0;
outArea = 0.0;
for (l=0; l<M-1; l++){
    if (concha->integrationBand[l].XM <=
        concha->integrationBand[l+1].XM) {
        topLimitIn = concha->integrationBand[l].XM;
        topLimitOut = concha->integrationBand[l+1].XM;
    } else if (concha->integrationBand[l].XM >
        concha->integrationBand[l+1].XM) {
        topLimitIn = concha->integrationBand[l+1].XM;
        topLimitOut = concha->integrationBand[l].XM;
    }
}
if (concha->integrationBand[l].xm <=
    concha->integrationBand[l+1].xm) {
    lowLimitIn = concha->integrationBand[l+1].xm;
    lowLimitOut = concha->integrationBand[l].xm;
}
else if (concha->integrationBand[l].xm >
    concha->integrationBand[l+1].xm) {
    lowLimitIn = concha->integrationBand[l].xm;
    lowLimitOut = concha->integrationBand[l+1].xm;
}
inArea = inArea + (topLimitIn - lowLimitIn) * concha->widthOfBands;
outArea = outArea + (topLimitOut - lowLimitOut) * concha->widthOfBands;
concha->area = (inArea + outArea) / 2;
printf("inArea = %f\t outArea = %f\t area = %f\n", inArea, outArea,
    concha->area);
} /* AnalysisConchaShape() */

/***************************************************************************/
/* fmax *****************************/
/***************************************************************************/
/* * This function finds the maximum value between two numbers. */
/**/
double fmax (double a, double b) {
    if (a > b) return (a);
    else return (b);
}

/***************************************************************************/
/* Find_xMax *****************************/
/***************************************************************************/
/* * This function finds the maximum x-coordinate amongst the read-in * concha shape values. */
/**/
void Find_xMax (double *xMax, XYDataPtr data) {
    int m;
    double tempMax;
    tempMax = 0.0;
    m = 0;
    while (m < data->numPoints) {
        tempMax = fmax(data->x[m], tempMax);
        m++;
    }
    *xMax = tempMax;
}

/***************************************************************************/
/* main function *****************************/
/***************************************************************************/

int main()
{
Appendix 3: Computer code

```c
XYDataPtr data = NULL;
char outFileFileName[35];
int j, k;
double elevation, azimuth, freq, coeff, degElev, degAzim, kK, wavelength, concha_xMax;
Complex pres, transfunct, T, reflectCoeff, diffPressrq, refPressrq;
Vector rq, rq_l;
FILE *fp;
Concha concha;

/* Read, Check and Print parameters --------------------------------*/
ReadParsFromFile (PARAMETERS_FILE);
CheckPars();
PrintPars();

/* Read concha contour (X,Y file) --------------------------------------*/
if (!Read_XYData(inputShapeFile, &data)) {
    fprintf (stderr, "Could not read file.\n");
    exit(1);
}

/* Analysis of the concha shape ----------------------------------------*/
Find_xMax (&concha_xMax, data);
printf ("Maximum X coordinate = %f, concha_xMax); AnalysisConchaShape(&concha, data);

reflectCoeff.re = reflectCoeff_Re;
reflectCoeff.im = reflectCoeff_Im;
for (degAzim = minAzimuth; degAzim < maxAzimuth + azimStep;
    degAzim += azimStep) {
for (degElev = minElevation; degElev < maxElevation + elevStep;
    degElev += elevStep) {
    /* create a file name according to the source location ----------*/
    printf ("\tElevation = %.If\n", degElev);
    printf ("\tAzimuth = %.If\n", degAzim);
    sprintf (outFileFileName,'%s_a%.Of%.Of[R%.lf].res', outputFile,
            degAzim, degElev, reflectCoeff.re);
    if ((fp = fopen(outFileFileName,"w+")) == NULL){
        printf ("Can't open %s file\n", outFileFileName);
        exit(1);
    }

    /* Printing parameters to file -----------------------------------*/
    PrintParsToFile (fp, degElev, degAzim);
    printf("Output being sent to %s file...\n", outFileFileName);
    elevation = degElev * (PI / 180.0); /* elev. angle to radians */
    azimuth = degAzim * (PI /180.0); /* azim. angle to radians */

    /* Evaluating the diffraction/reflection model for each freq */
    for (freq = minFreq; freq < maxFreq + freqStep; freq += freqStep){
        wavelength = SPEED / freq;
        kK = (2.0 * PI) / wavelength; /* kK is the wavenumber */
        for (j = 1, pres.re = 0.0, pres.im = 0.0;
            j< data->numPoints-1; j++) { /* open integration */
            rq.x = 0.0;
            rq.y = data->x[j];
```
Appendix 3: Computer code

rq.z = data->y[j];

rq_1.x = 0.0; /* rq_1 is in fact to r(q+1) */
rq_1.y = data->x[j+1];
rq_1.z = data->y[j+1];

if (rq.y == 0.0) {
    refPressrq.re = 0.0;
    refPressrq.im = 0.0;
}
else if (rq.y > 0.0) {
    diffPressrq.re = 0.0;
    diffPressrq.im = 0.0;

    /* Calculate total diffracted pressure over the reflecting 
     * element rq */
    for (k = 1; k < numBands - 1; k++) {
        DiffractionByBand(&T, k, rq, elevation, 
                           azimuth, KK, &concha);
        Add_CmplxM(&diffPressrq, &T, &diffPressrq);
    }

    /* Calculates the total reflected pressure */
    ReflectionBySurface(&refPressrq, rq, rq_1, 
                        &diffPressrq, KK, concha_xMax);
}

pres.re = pres.re +
    OpenIntegration_Coeff(j, data->numPoints) * refPressrq.re;
pres.im = pres.im +
    OpenIntegration_Coeff(j, data->numPoints) * refPressrq.im;

/* Multiplies the total reflected pressure by a kind of 
 * reflection coefficient.
 */
Mult_CmplxM(&pres, &reflectCoeff, &pres);
pres.re = pres.re + 1;
pres.im = pres.im;

/* Calculates the modulus of the transfer function */
Convert_CmplxM(&pres, &tranfFunct);
fprintf(fp, "%f\t%f\n", freq, 20 * log10(tranfFunct.re));
printf ("\nOutput sent to \$s file!\n", outFileName);
fclose (fp);
} /* elevation */
} /* azimuth */
} /* main */

*************************************************************************** End of 3DConcha10.c ***************************************************************************
Parameter file: 3DConcha10.par

This is the parameter file employed for obtaining the results shown in Figs. 5.4 and 5.5 (i.e., results for the realistic concha shape)

3D10 Output file name (used to number the output files for different conditions).
inputLateralConcha.dat Input file shape of the concha.
2.0 Number of diffracting bands.
1.0 Number of surface differentials within each band.
2.2 Concha Tilt (in degrees, between 0 and 90 degrees).
1.33 Concha Depth (in cm.).
1.0 Real part of the reflection coefficient.
0.0 Imaginary part of the reflection coefficient.
1.0 Minimum Frequency (Hz).
17000.0 Maximum Frequency (Hz).
250.0 Frequency Step (Hz).
0.0 Minimum Elevation (degrees, between -50 and +90).
0.0 Maximum Elevation (degrees, between -50 and +90).
10.0 Elevation Step (degrees).
0.0 Minimum Azimuth (degrees, between 0 and +90).
0.0 Maximum Azimuth (degrees, between 0 and +90).
10.0 Azimuth Step (degrees).

The file inputLateralConcha.dat' contained data pairs that define the YZ contour (Fig. 5.3a) of the realistic concha shape (see also Fig. A3.2).
APPENDIX 4

Microphones specifications

The manufacturer's frequency response curves for the Etymotic ER-11 and the Neumann KM-84i microphones employed by Gardner and Martin (1995) are shown below¹ (see Chapter 4).

The Etymotic ER-11 frequency responses are factory measurements of the actual microphones that Gardner and Martin had in their KEMAR while making the head-related impulse response measurements. They have a flat frequency response (± 0.5 dB) from 200 Hz to around 12 kHz. The response drops monotonically around 6 dB from 12 to 17 kHz.

The Neumann KM-84i frequency response has been taken from the manufacturer specification sheet. The dashed (thin) lines represent the maximum allowable deviation from the ideal response shown as the bold line, thus the KM-84i has a flat frequency response (± 2 dB) from 0.1 to 17 kHz.

The ER-11 microphones were used by Gardner and Martin to make the KEMAR HRIRs, whilst the KM-84i were used to measure the speaker response. Because the frequency response of both microphones is almost flat in the range 0.2 to 12 kHz, the equalisation procedure followed in Chapter 4 for compensating for the speaker response is therefore reasonable within that range. For frequencies in the range from 12 to 17 kHz, however, the head-related transfer functions shown in Chapter 4 must be around 6 dB lower than their actual value (i.e., than the value they would have if the speaker response had been measured with the ER-11 microphones).

¹ These curves were provided by B. Gardner.
ETYMOTIC RESEARCH
Elk Grove Village, Illinois 60007

305

FREQUENCY RESPONSE

Sensitivity

84dB SPL SOUND FIELD (0.316 Pascal)

6000 10000 20000
20 40 60

FREE IN HP

OUTPUT in dB re 1uV

860 1897 119
equency Responses and Polar Patterns

KM 831

KM 841

KM 851

KM 841