J. Acoust. Soc. Jpn. (*E*) **21**, 4 (2000)

REVIEW

Auditory images: How complex sounds are represented in the auditory system

Roy D. Patterson

Centre for the Neural Basis of Hearing, Physiology Department, Cambridge University, Downing Site, Cambridge, CB2 3EG, United Kingdom

When an event occurs in the world around us, say a lion roars or a cat meows, information about the event flows to us in light waves and sound waves. Our eyes form a visual image of the event, our ears form an auditory image of the event, and the two are combined with any other sensory inputs to produce our initial experience of the event. The purpose of this paper is to describe: 1) the three natural categories of sounds: noises, transients, and tones, 2) a computer model designed to explain how the auditory system converts sounds into auditory images, 3) the characteristics of the auditory images of noises, transients and tones, and 4) the role of auditory image construction in the initial segregation of tones and noises.

Keywords: Hearing, Auditory models, Auditory images, Temporal integration

PACS number: 43.64. Bt, 43.66. Ba, 43.66. Mk

1. INTRODUCTION

Sounds in the world around us fall broadly into three categories: noises, transients, and tones. The noises are generated mainly by moving air and water, like the sighing of the wind in the pines, and the washing of waves on the beach. In this environment many animals produce distinctive transients; shrimp and fish produce isolated transients to communicate, and cetaceans and bats produce transients to survey their environments. Most mammals also produce regular streams of transients that we hear as tones in their songs and calls. Humans produce both isolated transients and streams of transients. For example, the word 'at' consists of a stream of filtered glottal pulses (the vowel) followed by a band-pass filtered click (the plosive consonant). The first section of this paper describes a computational model designed to simulate how the auditory system converts a complex sound into an auditory image. The second section describes the properties of complex sounds as they appear in these simulated auditory images.

2. THE CONVERSION OF COMPLEX SOUNDS INTO AUDITORY IMAGES

The computational version of the auditory image model (AIM) is composed of a *cochlea simulation* that transforms a sound into a multi-channel neural activity pattern (NAP) like that observed in the auditory nerve,[1,2] and a form of *strobed temporal integration* that converts the simulated NAP into a representation of the *auditory image* we hear when presented with that sound.[3,4]†

2.1 Conversion of a Sound into a NAP

The cochlea simulation is composed of two processing modules: a *gammatone auditory filterbank* which performs the spectral analysis and converts the acoustic wave into a simulation of basilar membrane motion, and a bank of *two-dimensional adaptive-thresholding* units that *'transduces'* the

† AIM was originally set out in three papers written for conferences on auditory physiology (Keele), hearing (Carcans) and speech (Utrecht) in 1991. The papers appeared eventually as Refs. 1–3). AIM (Release 8) is distributed on the internet (aimR 8.2), and described in Ref. 4).

membrane motion and converts it into a representation of the multi-channel, neural activity pattern in the auditory system at the level of primary-like cells in the cochlear nucleus.

The gammatone auditory filter is defined in the time domain by its impulse response.

$$gt(t)=at^{(n-1)}\exp(-2\pi bt)\cos(2\pi f_c t+\phi) \quad (t>0) \tag{1}$$

It was introduced in Ref. 5) and validated Refs. 6,7) with 'revcor' data from cats. The primary parameters of the filter are b and n: b largely determines the duration of the impulse response and, thus, the bandwidth of the filter; n is the order of the filter and it largely determines the slope of the skirts. When the order of the filter is in the range 3-5, the shape of the magnitude characteristic of the gammatone filter is very similar to that of the rounded exponential or 'roex' filter commonly used to represent the magnitude characteristic of the human auditory filter.[8-10] The function relating the Equivalent Rectangular Bandwidth (ERB) of the auditory filter to its centre frequency (f_c) is

$$ERB=24.7+0.108f_c \quad \text{Hz.}^{11)} \tag{2}$$

So, to a first approximation, the bandwidth is 25 Hz plus 10 percent of the centre frequency. This function is essentially the same as the 'cochlear frequency position' function.[12] which is the physiological basis for the 'critical band' function. Together Equations 1 and 2 define a gammatone auditory filterbank if one includes the common assumption that the filter centre frequencies are distributed across frequency in proportion to their bandwidth. When f_c/b is large, as it is in the auditory case, the bandwidth of the filter is proportional to b, and the proportionality constant only depends on the filter order, n. When the order is 4, b is $1.019 \cdot ERB$. The 3-dB bandwidth of the gammatone filter is $0.887 \cdot ERB$. Examples of the response of the gammatone filterbank to vowels and other complex sounds are presented in Refs. 1-4).

The auditory filter concept has been extended[13] to include the asymmetry in auditory masking observed at high stimulus levels.[10] This level-dependent, gammachirp auditory filter can also explain the details of the impulse responses derived[14] directly from physiological (revcor) data.[15] For a recent review of peripheral filtering see Ref. 20).

Neural transduction is performed in the cochlea by 'hair cells' which convert the motion of the basilar membrane into neural transmitter. The hair cells compress membrane motion adapting to level changes rapidly. Primary-like cells in the cochlear nucleus apply lateral inhibition across frequency. Together these processes enhance features that arise in basilar membrane motion, indicating that this neural processing should be regarded as a sophisticated signal processing system designed to emphasise spectro-temporal features of the incoming sound.

In AIM, the conversion of basilar membrane motion into neural activity is simulated with a cascade of four, multi-channel signal-processing modules: a bank of half-wave rectifiers, a bank of logarithmic compressors, a bank of adaptation units, and a bank of low-pass filters to simulate the progressive loss of phase locking at frequencies above 1,200 Hz. There is one rectifier, compressor, adaptation unit, and low-pass filter for each channel of the auditory filterbank. The central function in Fig. 1 shows the simulated neural activity produced by a 2,000-Hz sinusoid, amplitude modulated by a 120-Hz sinusoid. It shows that the adaptation is asymmetric in time. Upwards adaptation to onsets is virtually instantaneous; recovery from this upwards adaptation is much slower. Examples of the response of the full cochlea simulation to temporally-asymmetric complex sounds are presented in Refs. 16-19). In the frequency dimension, the adaptation is asymmetric on a linear frequency axis, and nearly symmetric on a log-frequency axis. The name two-dimensional adaptive-thresholding (2D-AT) emphasises the fact that adaptation is applied in the frequency dimension as well as the time dimension.[1] The module converts AIM's representation of basilar membrane motion into another surface that represents AIM's approximation to the multi-channel NAP at the level of the cochlear nucleus. In summary, 2D-AT is a functional representation of neural processing in which a cascade of four signal-processing modules simulate the aggregate activity of all of the inner hair cells and primary fibers associated with one auditory filter (about 0.9 mm of basilar membrane in humans).

2.2 The NAP as a Representation of What We Hear in a Sound

It is often assumed that peripheral auditory

processing ends at the output of the cochlea and that the pattern of activity in the auditory nerve or cochlear nucleus is what we hear. The multi-channel NAP produced by a full cochlea simulation is a better representation of what we hear in a sound than the output of a linear auditory filterbank.[16] However, there are several problems with the NAP as a representation of our auditory image of a sound. To begin with, there is *between-channel phase information* in the NAP which we do not hear. For example, if we compensate for the cochlear phase delay, we can synchronise firing in all frequency channels. But this does not affect what we hear. It is phase changes within channels that lead to changes in sound quality.[21]

Another problem with NAPs is the contrast between the *perceptual stability of tonal sounds* and the detail we hear in the timbre of these sounds. In response to a vowel sound, the level of activity in the NAP varies from maximum to minimum over the course of the glottal cycle, say 8 ms, but we do not hear the loudness of the sound varying from deafening to silent and back 125 times per second. Periodic sounds, like sustained musical notes and sustained vowels, produce static auditory images with fixed loudness, fixed pitch, and fixed timbre. This indicates that some form of fairly lengthy temporal integration is applied to the NAP during the production of our initial perception of a sound —our auditory image. For some reason, however, the temporal integration does not seem to destroy the fine temporal detail, when the sound is periodic or quasi-periodic ; we hear the most delicate timbres in static sounds. This is a puzzle because auditory temporal integration has traditionally been modelled with leaky integrators which average over time and so destroy temporal fine-structure.[22]

The NAP in Fig. 1 includes simulation of the loss of phase locking that occurs during neural transduction ; it is implemented as a two-stage low-pass filter with a time constant of 0.133 ms (a cut-off frequency of 1,200 Hz). Nevertheless, there is still considerable phase-locking information evident in this channel. At the same time, the NAP oscillates far too fast to explain the fixed perception that this sound produces. We could add another leaky integrator to stabilise the representation, like that used to explain the temporal-modulation transfer function.[22] But the time constant in that model is 2.5-ms—more than ten times that associated with the

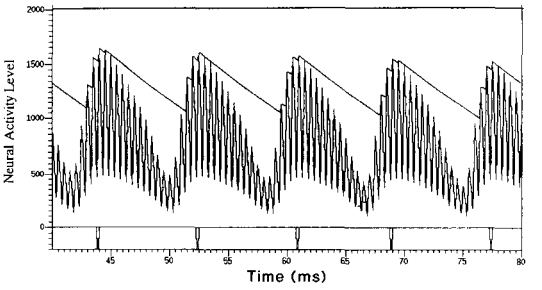

Fig. 1 A 40-ms segment of the NAP produced by an amplitude modulated sinusoid ; the carrier is a 2,000-Hz sinusoid and the modulator is a 120-Hz sinusoid. Above the NAP is the adaptive threshold of the strobe unit; below the NAP are the strobe points where temporal integration is initiated.

NAP in Fig. 1. It greatly reduces the phase-locking information but the output still oscillates too fast to explain the stability of the perception. Another option is the temporal window used to explain the detection of tones presented in gaps in Ref. 23) ; it is implemented as a two-stage low-pass filter with a 5-ms time constant. It increases the stability but it eliminates temporal fine structure entirely. So, integration times long enough to produce the stability we hear, destroy the temporal fine structure observed in the auditory nerve—information that is necessary to explain pitch perception,[24,25] sound quality[26-28] and phase perception.[21]

2.3 Conversion of the NAP into a Stabilised Auditory Image (SAI)

In AIM the stability problem is solved with '*strobed*' temporal integration.[3,17,19,26] It is assumed that the auditory system has a bank of delay lines to form a buffer store for the NAP as it flows from the cochlea, and a bank of time-interval histograms into which the system deposits time intervals measured between pulses in the NAP. Each channel of the NAP is like a chart recording flowing from a cochlear channel positioned at the right-hand edge of Fig. 1. The auditory image is simulated with a static image buffer (Fig. 2). Once information about past events is integrated into the image, the information does not change position ; it just decays in place as time passes. The selection of the time intervals that proceed from the NAP to the interval histogram is controlled in each channel by a 'strobe' unit whose operation is illustrated by the

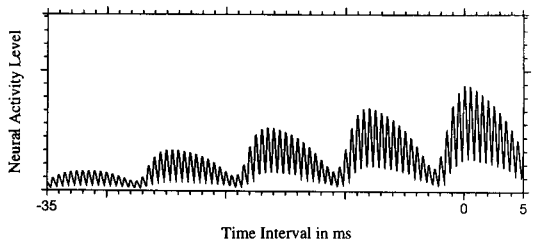

Fig. 2 The 2,000-Hz channel of the auditory image produced from the NAP of the amplitude modulated sinusoid of Fig. 1.

jagged envelope on top of the NAP in Fig. 1. AIM's simulation of the resultant interval histogram is shown in Fig. 2. The strobe mechanism is an adaptive threshold that monitors the activity in its channel looking for a local maximum in NAP activity. After each peak the adaptive threshold decays relatively slowly (about 5%/ms) and when it decays for 5 ms without encountering a larger peak in the NAP, the previous peak is designated a strobe pulse, and the unit initiates Strobed Temporal Integration (STI) in that channel.

In the auditory system, STI would involve calculating time intervals from the strobe pulse to all of the other pulses currently in that channel of the NAP, and then, for each time interval, incrementing the corresponding point in the interval histogram by an amount proportional to height of the strobe pulse. As the NAP proceeds down the delay lines of the buffer store, the level of the pulses in the buffer is assumed to decay linearly with time at about 2.5 %/ms. This decay is not shown in Fig. 1; in the computational model it is applied as the time-interval histogram is constructed. The decay appears in the interval histogram (Fig. 2); the decay rate was set so that there is no activity in the buffer beyond 40 ms, since the lower limit of pitch for complex sounds is just above 25 Hz.[29] This limit was used[30] to explain the perception of jitter in click trains and the fact that sensitivity to jitter decreases as the period of the click train increases beyond 30 ms.

In the auditory system, the multi-channel interval histogram is assumed to be the basis of our auditory images. In the computational version of AIM, STI is initiated on each strobe and it is implemented simply by adding a copy of the NAP at that moment to the activity in the corresponding channel of the auditory image buffer. The time associated with the peak of the NAP pulse that produced the strobe is mapped to the 0-ms point in the interval histogram. The STI process converts the *time* dimension of the NAP (Fig. 1) into a *time-interval* dimension in the image (Fig. 2). As a result, once a contribution from the NAP arrives in the image it does not move left with time any longer; activity in the image simply decays away exponentially into the floor. The half life of the auditory image is assumed to be about 30 ms, and so individual transients disappear promptly. For periodic and quasi-periodic sounds, the STI mechanism rapidly adapts to the period of the sound and initiates temporal integration roughly once per period of the sound. In this way, STI matches the temporal integration period to the period of the sound and, much like a stroboscope, it produces a static auditory image of the repeating temporal pattern in the NAP as long as the sound is stationary. The image of a periodic sound grows rapidly over the first 4–5 cycles. Thereafter, the level rises more slowly and the image stabilises as the integration and decay processes come into balance. When a sound changes abruptly from one form to another, the auditory image of the initial sound simply collapses and is replaced by the image of the new sound. If the rate of change is slow relative to the rate of repetition of the pattern in the NAP, as in the case of diphthongs in speech and gliding notes in music, then the pattern in the auditory image changes smoothly from one state to another in much the same way as characters move smoothly in an animated cartoon.

3. THE CHARACTERISTICS OF NOISES, TRANSIENTS AND TONES IN THE AUDITORY IMAGE

Simple flat-spectrum examples of the three categories of sound are white noise, an acoustic impulse, and a click train. They exercise the full range of frequency channels and illustrate the patterns that are characteristic of the three categories of sounds as they appear in the auditory image (Fig. 3). The image produced by the click train (top) reveals a vertical structure which is referred to as an auditory figure. In this case, it is the neural version of the multi-channel impulse-response produce by a click in the gammatone auditory filterbank. The figure repeats at regular intervals across the image, and the interior of the auditory figure is highly regular.

The presence of pattern across the image is the characteristic of tonal sounds—regularity, on both the macro and the micro scale. The *horizontal spacing* of the auditory figures reveals the *pitch* of the sound ; the spacing between auditory figures decreases as pitch increases. The lower limit of pitch is about 30 Hz, corresponding to a period of 33 ms, which is the maximum width of the auditory image.[29] The image produced by white noise (middle) reveals activity across the full width of the image, but in this case, there are no auditory figures and there is no repetition of the occasional random feature that does arise in the noise. The absence of pattern is the characteristic of noisy sounds, both on the macro and micro scale. The regularity about the 0-ms vertical in the noise image is an artefact of strobed temporal integration ; strobing is always initiated on a NAP peak and the peak is always mapped to 0-ms in the auditory image. The isolated click (transient) generates an image (bottom) that consists of one auditory figure centred on the 0-ms line, and no activity in the remainder of the image. The auditory figures of transients fade out of the image rapidly because the half life of the image is just 30 ms. As a result, the auditory system often misses the details of the transient when it occurs without warning. But if it repeats, as when a horse walks slowly down a cobbled street, the individual shoe figures become sufficiently distinctive to tell us, for example, if one of the shoes is loose or missing. In summary, the temporal and spatial properties observed in these three auditory images are consistent, distinguishing characteristics of the auditory images of tones, noises and transients.

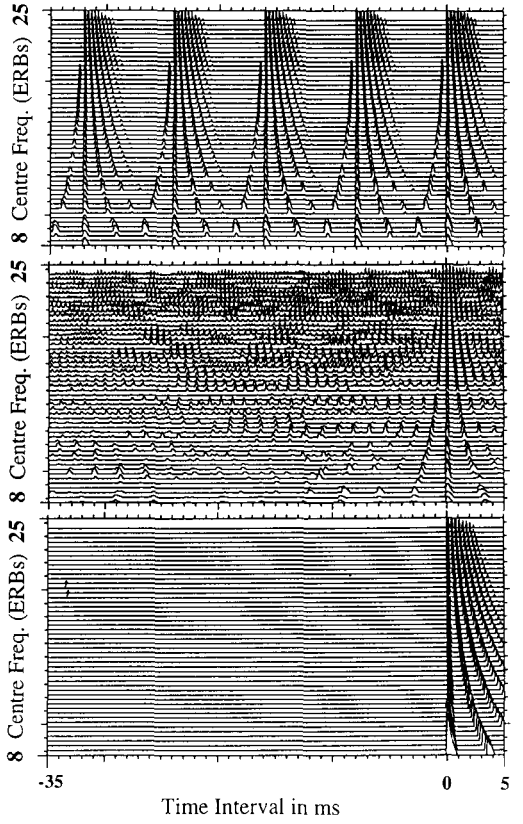

Fig. 3 Auditory images of a click train (top), a white noise (middle), and a single click (bottom). The abscissa is Time Interval and the ordinate is the centre frequency of the channel.

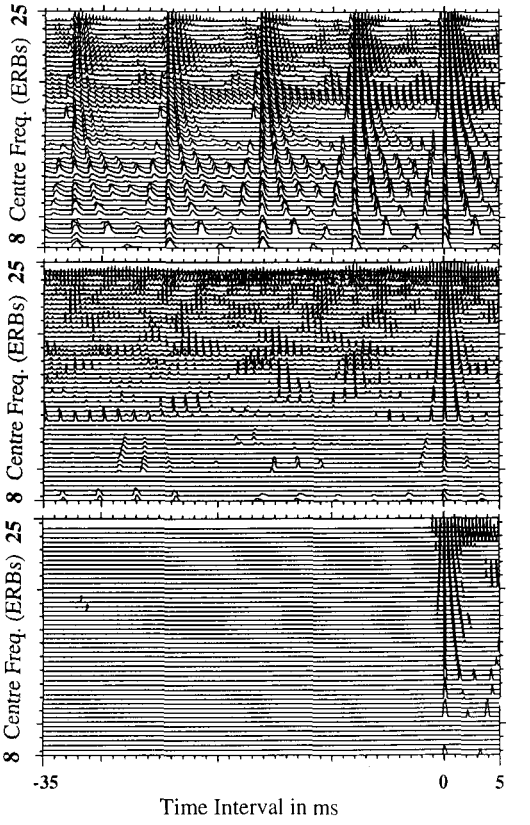

Fig. 4 Auditory images of the 'a' (top), 's' (middle), and 'k' (bottom) of the word 'ask'. The abscissa is Time Interval and the ordinate is the centre frequency of the channel.

3.1 Auditory Images of Real-World Sounds

Many real-world sounds are combinations of tones, noises and transients, including speech sounds. Auditory images of *the three phonemes in the word 'ask'* are presented in Fig. 4. The /a/ is a speech tone, or vowel; the /s/ is a speech noise, or fricative consonant; and the /k/ is a speech transient, or plosive consonant. The distribution of activity in the vertical dimension is less uniform than for the prototypical sounds of Fig. 3, but the similarity between the auditory images of the components of 'ask' and the prototypical sounds is clear. The auditory figures in the images of the click train (Fig. 3) and the vowel (Fig. 4) have approximately the same horizontal spacing and the sounds have approximately the same pitch (125 Hz). The auditory figures of the click train and vowel differ inasmuch as the auditory figure of the vowel has a more complex *shape* and it has a rougher *texture*. The shape and texture of the simulated auditory images capture much of the character of the 'sound quality' or 'timbre' of the sound. The *shape* of the vowel figure is largely determined by the centre frequencies and bandwidths of the resonances in the vocal tract of the speaker at the moment of speaking. The formants produce the horizontal features in the auditory figure and they move up and down as the vowel changes. So different shapes correspond to different vowels.[2,18] *Texture* is primarily determined by the degree of periodicity of the source in the individual channels of the image. The texture of the formants in the auditory figure—the degree of definition in the simulated image—provides information concerning whether the speaker has a breathy voice, whether the consonants in the syllable are voiced, and whether the syllable is stressed or not.

The differences between the auditory images of the white noise and /s/ are largely spectral; the white noise has relatively more low-frequency energy. Similarly, the main difference between the sound of wind in the trees and the sound of a waterfall is that the latter has a higher proportion of low frequency energy. The main difference between the transients that we hear as the stop consonants /p/ and /t/ is that the /p/ has much more low frequency energy. In general, however, transients are more complex and the transient excitation is accompanied by a secondary transient (as in the flapped /d/ in the American pronunciation of 'water') or by a segment of noise (as with /p/'s in initial position).

3.2 Temporal Regularity and Figure/Ground Separation in the Auditory Image

During the construction of the auditory image of a periodic sound, the strobe mechanism synchronises temporal integration so that pulses in the microstructure of the NAP add up to form peaks while the small gaps between them remain empty. With noise, the effect is largely the reverse. Each channel of the NAP is generated by a narrowband noise from one auditory filter, so it contains a stream of NAP pulses separated by gaps just as for a periodic sound. The average spacing of the NAP pulses reflects the centre frequency of the filter, but the time between pulses varies about the period of the centre frequency. In the short term, the relatively narrow bandwidth of the auditory filter limits the rate of change, and this produces the regularity observed in the region close to 0 ms in the auditory image of a noise. In the longer term, however, the position of NAP pulses in time varies randomly with regard to the period of the centre frequency. As a result, the time between strobes is random and the time-interval values are randomly distributed. So, in the auditory image, noisy sounds produce asynchronous temporal integration, and away from the origin, the pulses are distributed evenly across time interval. Thus, NAP pulses do not add up and, as often as not, they fall in the gaps between previous sets of pulses, with the result that the level averages down towards the mean value.

The effect of strobed temporal integration on mixtures of periodic and aperiodic sounds is like the triggered averaging used to extract weak physiological signals from noisy measurements. The strobe process actually enhances the signal-to-noise ratio of any periodic components in the mixture. At the same time, the process destroys some of the information in the noise, but this is not a major source of information in environmental sounds.

3.3 Auditory Images Simulated with Autocorrelation

AIM is unique in its emphasis on strobed temporal integration for the production of auditory images from activity in the auditory nerve. It should be noted, however, that displays similar to the SAI were produced as early as 1984[31] from the

first computational version of the Duplex Theory of Pitch Perception.[32] A running autocorrelation was performed on the output of each channel of a cochlea simulation, and the result was plotted as a grey scale display with autocorrelation lag on the abscissa and channel centre frequency on the ordinate. Although the motivation for implementing the autocorrelation mechanism was pitch extraction, it was pointed out that the resulting 'autocorrelogram' contains information about the position of the formants as well as the pitch. More recent implementations of the Duplex Pitch Model[33-36] confirm the value of time-interval processing for identification of vowels in noise. It should be noted, however, that autocorrelation is a symmetric process in time and, as a result, autocorrelogram models of perception have difficulty explaining listener's ability to discriminate pairs of sounds that differ only in their temporal asymmetry.[17,19]

4. CONCLUSION

Time-domain models of the early stages of auditory processing can now provide a bridge between the output of the cochlea as observed in single unit studies with animals and the auditory images that we hear when presented with complex sounds like speech and music. In the Auditory Image Model, the multi-channel output of the cochlea is simulated with a gammatone auditory filterbank followed by a bank of adaptive thresholding units. Then a form of strobed temporal integration is used to stabilise repeating neural patterns in a multi-channel image buffer which simulates the auditory images that we hear.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by the MRC and DERA (project AAM HAP) of the UK, a grant from the European Esprit BRA (project ACTS), and a grant ATR/HIP in Kyoto.

REFERENCES

1) R. D. Patterson and J. Holdsworth, "A functional model of neural activity patterns and auditory images," in *Advances in Speech, Hearing and Language Processing*, W. A. Ainsworth, Ed., Vol. 3, Part B (JAI Press, London, 1996), pp. 547-563.

2) R. D. Patterson, J. Holdsworth, and M. Allerhand, "Auditory Models as preprocessors for speech recognition," in *The Auditory Processing of Speech: From the Auditory Periphery to Words*, M. E. H. Schouten, Ed. (Mouton de Gruyter, Berlin, 1992), pp. 67-83.

3) R. D. Patterson, K. Robinson, J. W. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex sounds and auditory images," in *Auditory Physiology and Perception*, Y. Cazals, L. Demany, and K. Horner, Eds. (Pergamon, Oxford, 1992), pp. 429-446.

4) R. D. Patterson, M. Allerhand, and C. Giguere, "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890-1894 (1995).

5) E. de Boer, "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030-1045 (1975).

6) E. de Boer and H. R. de Jongh, "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique," J. Acoust. Soc. Am. **63** 115-135 (1978).

7) L. Carney and C. Yin, "Temporal coding of resonances by low-frequency auditory nerve fibers: Single fibre responses and a population model," J. Neurophysiol. **60**, 1653-1677 (1988).

8) R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788-1803 (1982).

9) R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, B. C. J. Moore, Ed. (Academic Press Limited, London, 1986), pp. 123-177.

10) S. Rosen and R. J. Baker, "Characterising auditory filter nonlinearity," Hear. Res. **73**, 231-243 (1994).

11) B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138 (1990).

12) D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592-2605 (1990).

13) T. Irino and R. D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412-419 (1997).

14) L. H. Carney, J. M. Megean, and I. Shekhter, "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384-2391 (1999).

15) T. Irino and R. D. Patterson, "A gammachirp perspective on cochlear mechanics that can also explain auditory masking quantitatively," Proc. Auditory Mechanics Workshop, Sendai, Japan (1999).

16) R. D. Patterson, "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409-1418 (1994).

17) T. Irino and R. D. Patterson, "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **99**, 2316-2331 (1996).

18) T. Irino and R. D. Patterson, "Stabilised wavelet mellin transform: An auditory strategy for normalis-

ing sound-source size," Eurospeech 99, Budapest, Sept (1999).
19) R. D. Patterson and T. Irino, "Modeling temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **104**, 2967-2979 (1998).
20) M. Akagi, "Sound wave analysis on auditory characteristics," J. Acoust. Soc. Jpn. **54**, 575-581 (1998).
21) R. D. Patterson, "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am. **82**, 1560-1586 (1987).
22) N. F. Viemeister, "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364-1380 (1979).
23) C. J. Plack and B. C. J. Moore, "Decrement detection in normal and impaired ears," J. Acoust. Soc. Am. **90**, 3069-3076 (1991).
24) W. A. Yost, R. D. Patterson, and S. Sheft, "A time-domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066-1078 (1996).
25) W. A. Yost, R. D. Patterson, and S. Sheft, "The role of the envelope in processing iterated rippled noise," J. Acoust. Soc. Am. **104**, 2349-2361 (1998).
26) R. D. Patterson, "The sound of a sinusoid: Time-interval models," J. Acoust. Soc. Am. **96**, 1419-1428 (1994).
27) R. D. Patterson, S. Handel, W. A. Yost, and A. J. Datta, "The relative strength of the tone and noise components in iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286-3294 (1996).
28) R. D. patterson, W. A. Yost, S. Handel, and A. J. Datta, "The perceptual tone/noise ratio of merged iterated rippled noises," J. Acoust. Soc. Am. **107**, 1578-1588 (2000).
29) K. Krumbholz, R. D. Patterson, and D. Pressnitzer, "Period difference limens for harmonic complex tones in and below the pitch region," in *Psychophysics, Physiology and Models of Hearing*, T. Dau, V. Hohmann, B. Kollmeier, Eds. (World Scientific, Singapore, 1999).
30) M. Tsuzaki and R. D. Patterson, "Jitter detection: A brief review and some new experiments," in Psychophysical and physiological advances in hearing: Proc. 11th International Symposium on Hearing, A. Palmer, A. Rees, Q. Summerfield, and R. Meddis, Eds., Whurr, London, 546-553 (1998).
31) R. F. Lyon, "Computational models of neural auditory processing," Proc. ICASSP 84, San Diego, USA (1984).
32) J. C. R. Licklider, "A duplex theory of pitch perception," Experientia **7**, 128-133 (1951).
33) P. F. Assman and Q. Summerfield, "Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies, J. Acoust. Soc. Am. **88**, 680-697 (1990).
34) M. Slaney and R.F. Lyon, "A perceptual pitch detector," in Proc. ICASSP 90, Albuquerque, New Mexico (1990).
35) R. Meddis and M. J. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I pitch identification," J. Acoust. Soc. Am. **89**, 2866-2882 (1991).
36) T. D, Griffiths, C. Beuchel, R. S. J. Frackowiak, and R. D. Patterson, "Analysis of temporal structure in sound by the brain," Nat. Neurosc. **1**, 422-427 (1998).