Proceedings of The Institute of Acoustics

SPECTRAL ANALYSIS OF COMPLEX SOUNDS AT THE AUDITORY PERIPHERY

E.F. EVANS

DEPT. OF COMMUNICATION & NEUROSCIENCE, UNIVERSITY OF KEELE

INTRODUCTION

This paper outlines recent findings in our laboratory concerning the spectral analysis of complex sounds at the cochlear nerve level of the auditory system. Further accounts are to be found in (6, 3, 9, 11).

At the cochlear nerve level, the spectral analysis is surprisingly linear in spite of the non-linear mechanisms now known to be involved in determining cochlear frequency selectivity (see 10 for review). Thus, the concepts of effective bandwidth and filter shape have a real meaning in predicting the responses of neurones at this level to complex stimuli: broadband and comb-filtered noise; harmonic complexes; click complexes.

The cochlear nerve considered as a bank of filters. Each of the 30,000 fibres in the mammalian cochlear nerve can be considered, to a first approximation, to act as a narrow band filter. Each fibre has a preferred, centre frequency (the characteristic frequency, CF) reflecting its origin along the cochlear partition. Stimulus energy falling within the triangular response area bounded by the frequency threshold ('tuning') curve (FTC; see continuous curve 'tone threshold' in Fig. 1A) will evoke an acceleration of the mean rate of impulse activity of the fibre above its resting discharge of 0-120 spikes/s. The cut-offs and the bandwidths of the FTC depends upon the fibre's characteristic frequency. For fibres with CF above about 2kHz, the FTC is asymmetrical and the cut-offs steep, particularly on the high frequency side, approaching (in the cat and guinea pig) 500 dB/oct. For lower CFs, the curves are more symmetrical and the cut-offs become progressively less steep, from 50-100 dB/oct at 1kHz to about 10-20 dB/oct at 0.2kHz (3). The effective bandwidths of the cochlear fibre filters for broad-band signals are approximately their half-power bandwidths (16). They range from about 100 Hz for fibres with CFs up to about 1kHz, and 10% of the CF thereafter (Fig. 4; 16). Although there is systematic variation in these values from individual to individual (see the differences between cats in Fig. 4), they are representative of measurements of filtering employing pure tone, broad-band and comb-filtered noise stimuli (16), within limits to be described later.

This remarkably sharp filtering is also reflected in the temporal patterning of the discharges for frequencies up to about 4-5kHz (Figs. 1, 2, 3). In Fig. 1, the results of reverse correlation analysis with a broad-band noise stimulus are shown for a fibre in the cat's cochlear nerve having a CF of 2kHz and the FTC shown by the continuous curve in Fig. 1A (4). The reverse correlation analysis yields the impulse response of the cochlear filter, shown in Fig. 1B for various levels of the wideband noise stimulus. The Fourier transforms of the impulse responses are shown for comparison with the pure tone FTC. Several points may be noted. First, the correlation analysis indicates that the filtering represented by the pure tone FTC is reflected in the frequencies dominating

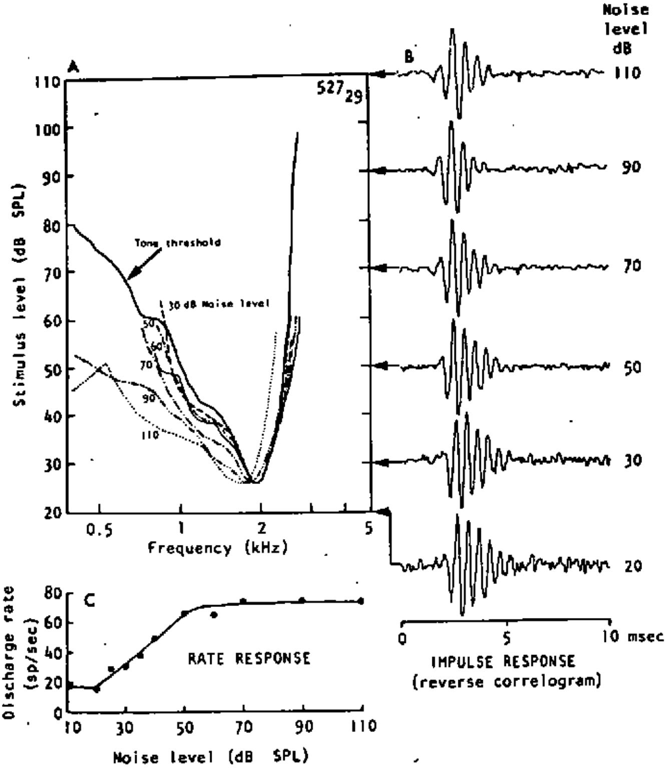

Fig. 1. Filtering characteristics of a cochlear fibre, derived by reverse correlation at different noise stimulus levels.
A: Frequency threshold ('tuning') curve (continuous line) with pure tone stimulation for comparison. Interrupted lines: Fourier transforms of the impulse responses (B) obtained by reverse correlation of the spike discharges with the broad-band noise stimulus. C: Discharge rate versus stimulus level function for the responses in A, B, showing the restricted dynamic range (30-40 dB) typical of the majority of cochlear fibres. (From 4)

the temporal patterns of discharge, on which the analysis depends. In other words, the discharge pattern tends to follow preferentially frequencies within the effective bandwidth of the fibre. Secondly, the impulse response is more like that of a multiple pole band-pass than a resonant filter, reflecting the U (rather than V) shape of the FTC tip. Thus, the cochlear filter appears to have effected an optimal compromise between spectral and temporal resolution, the

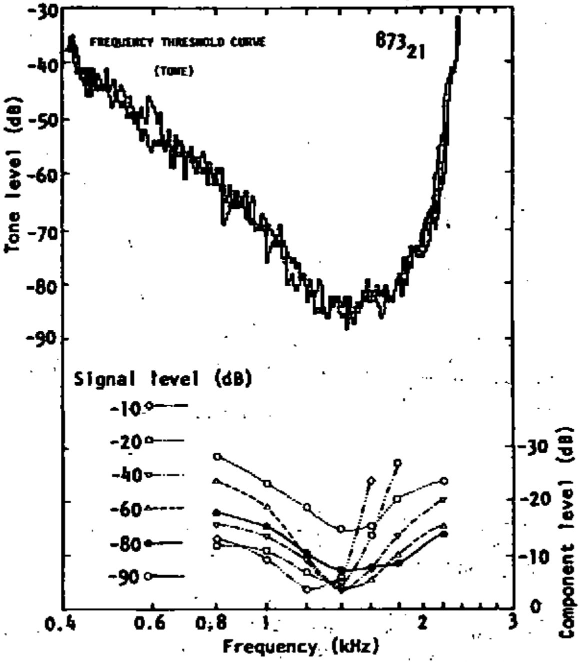

Fig. 2. Mean discharge rate threshold (Frequency Threshold Curve) of a single cochlear fibre compared with the vector strengths of its phase-locked response to individual harmonics of a ten-harmonic tone-complex at different stimulus levels. Vector strengths plotted on a log power scale for comparison. 0dB = 106 dB SPL. (From 13)

latter being of the order of 5ms for fibres with CFs in the region of 1kHz. Thirdly, the bandwidth of the cochlear filter increases and the low-frequency cut-off decreases with stimulus level, particularly above 30-40 dB above threshold. This does not happen in all fibres, at least in the cat; and in any case, the effective bandwidths increase by a factor of only 1.1-2.5 at the highest noise levels (100-110 dB SPL), again in the cat. The small but systematic shift in the CF with level is CF dependent: for fibres with CF above about 1kHz, it shifts downwards by about 10%, and for those below 1kHz, it shifts upwards (4).

The properties are also reflected in the temporal pattern of activity of cochlear fibres to stimuli having multiple frequency components (Fig. 2) (8, 13). The fibre in Fig. 2 is being stimulated with a 10 harmonic (equal amplitude) complex centred on its CF (1.4kHz), the fundamental frequency being 200 Hz. The degree to which the temporal patterns of discharges synchronize with each harmonic ('phase-locking') is plotted as the 'vector strength' (vd. 2, 17) against the

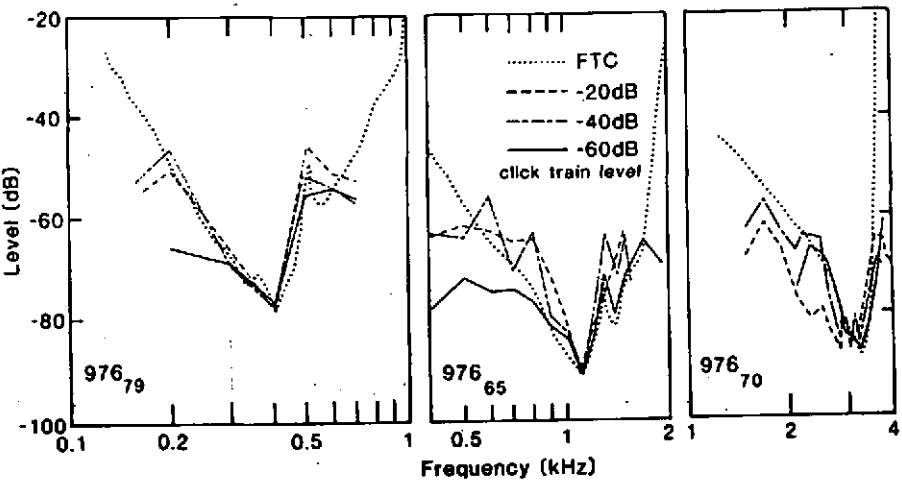

Fig. 3. Comparison of weighting functions derived from the Fourier transforms of autocorrelation histograms in response to the click train, for 3 fibres with CFs of 0.37, 1.08 and 3.15kHz respectively at 3 stimulus levels in each case, with the pure-tone FTC of each fibre.
FTCs indicated by dotted lines; weighting functions: -60 dB level (about 10 dB above response threshold): continuous line; -40 dB level: dashed-dotted line; -20 dB level (about 50 dB above threshold): dashed line. (From 11).

signal level. These vector strengths are plotted against frequency in Fig. 2, on a logarithmic scale so that comparison may be made with the pure tone FTC. Again, at stimulus levels near threshold, the frequency weighting of the vector strengths reflects the FTC. At higher stimulus levels, similar changes in selectivity are reflected in the vector strengths as described in the case of reverse correlation (13). Similar results have been obtained with impulsive stimuli (Fig. 3). This shows the weighting, by the cochlear nerve filtering process, of a fibre's response to a click train (11). It is derived by FFT of the autocorrelation histogram of the discharge pattern evoked by the click train. The figure shows the correspondence between the weighting functions at 3 levels of the click train, with the pure tone FTC (dotted outline). The minor deviations between the weighting functions and the FTC are consistent with those encountered with broad-band noise and harmonic complexes as described above. At threshold stimulus levels, the weighting function with harmonic complexes is sometimes slightly wider than the FTC. At stimulus levels up to about 40-60 dB above threshold, however, the correspondence is close, particularly for fibres with CFs below about 1.5kHz. For higher stimulus levels, the weighting function tends to shift towards lower frequencies for fibres with CFs above about 1kHz (Fig. 3, fibre 976_{70}) together with a progressive decrease in the slope of the low frequency cut-off.

To a first approximation, therefore, cochlear nerve fibres act for broad-band and multicomponent stimuli as if they were linear filters having the filter characteristics described by their pure tone FTC. Given a fibre's CF, its properties in response to complex sounds can therefore be reasonably well predicted, and simulated by simple electronic analogues (12). In other words, it is those frequencies that fall within the effective bandwidth of the fibre's filter, that tend to dominate its mean discharge rate, and the temporal pattern of the discharges. As has been indicated above, this is true for the latter over a very wide dynamic range of intensity.

In the case of mean discharge rate, however, the above approximation breaks down above levels of about 40 dB above threshold. Fig. 1C shows how the level of the stimulus can be signalled in the mean discharge rate of the fibre only over a range of about 30-40 dB, before it saturates. At stimulus levels above saturation, then, cochlear fibres cannot manifest their frequency selectivity in their discharge rate properties, although as indicated in Figs. 1 and 2 for those with CFs below about 3-5kHz (the limit of 'phase-locking') they can do so in the degree of synchrony of their temporal discharge patterns with the stimulus.

Cochlear fibre responses to complex stimuli, particularly two component stimuli, also deviate from the above first approximation description, in respect of certain cochlear non-linearities responsible for intermodulation distortion and 'lateral suppression' effects. Of the former, the $2f_1-f_2$ cubic distortion product is the most important (e.g. 18), although its presence is evident only in carefully contrived stimulus conditions. The latter non-linear effect, termed two-tone suppression or inhibition (e.g. 32; although it is not limited to two tones, nor is it neural inhibition) means that the excitation generated by a stimulus at one frequency (e.g. at the CF) can be suppressed by that at another, particularly, lower, frequency. This suppression is reflected in both the mean discharge rate response and in the degree of synchrony exhibited by the temporal discharge patterns. Thus, in the fibre of Fig. 2, the synchrony to harmonic frequencies at and above the CF (e.g. 1.6kHz) becomes progressively more suppressed with stimulus level, reflecting the increasing dominance of the frequency components below the CF. This phenomenon has been termed 'synchrony suppression' (31). For speech stimuli, these lateral suppression effects mean that the lower frequency, higher energy first formant will have suppressing effects on the rate and synchrony of the responses to the higher formants (36). It should be emphasised, however, that these effects are not strong features of the responses of the majority of cochlear fibres having spontaneous discharge rates above about 20 sp/s. The effects are strongest in the smaller subpopulation of the fibres with the lowest spontaneous discharge rates (33).

Physiological and psychophysical filtering

The above measurements of physiological cochlear filtering are relevant also because they are consistent with measurements of human frequency selectivity obtained by psychophysical methods. Thus, the physiological effective bandwidths (obtained in cats, see data points in Fig. 4) approximate to the values of the human 'critical bands' (dashed line in Fig. 4), becoming smaller than the latter for frequencies above 1kHz (16; see 5 for review). The shapes and cut-off slopes of the cochlear fibre FTCs are consistent with those obtained by tone-on-tone masking techniques yielding the so-called 'psychoacoustic tuning curves' (e.g. 35,

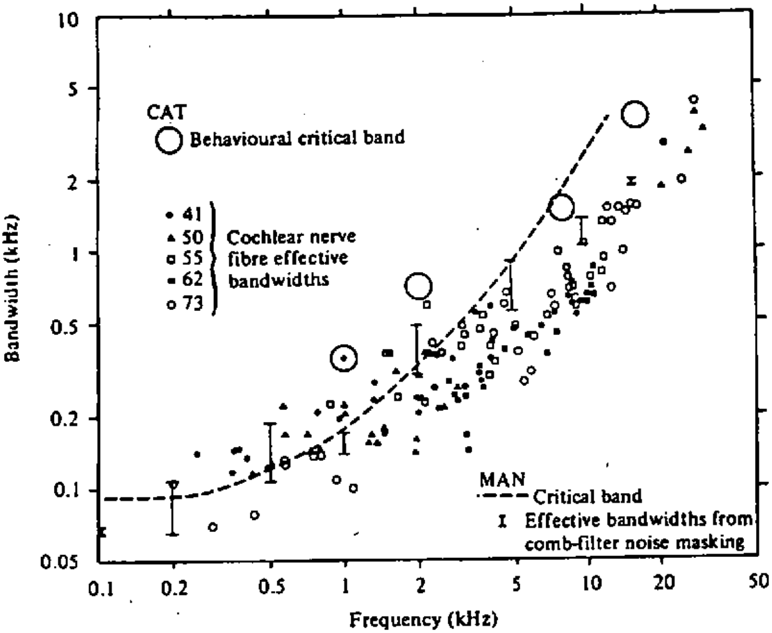

Fig. 4. Comparison of neural and psychophysical effective filter bandwidths. Each small symbol represents the effective bandwidth of an individual cochlear nerve fibre filter plotted against its characteristic frequency. Individual cats are identified by different symbols. Dashed line represents the 'critical band' of human hearing, one estimate of the bandwidth of human auditory filters determined psychophysically. Human effective bandwidths determined by comb-filtered noise masking are also included. Open circles: behavioural measurements of 'critical band' in the cat (data from 27). (From 14).

37). It is not possible to reconcile entirely the filter bandwidths derived from all the different psychoacoustic (e.g. 21, 24) and animal behavioural (e.g. 27, 28) techniques of measurement in order to compare them with the physiological effective bandwidths. Taken together, they suggest that a value of about 10% (i.e. about 1/6 octave) would be more appropriate for the effective bandwidth of the ear's peripheral filter bank than the value of 1/3 octave commonly employed in modelling the 'front-end' of the auditory system, and representing the larger critical band value obtained in simultaneous masking.

This limited frequency resolution imparts important constraints on the first stage of the processing of complex stimuli by the auditory system. It sets limits on the closeness of individual frequency components that can be resolved; and conversely, it determines which components of a complex sound will interact so that the gross and fine temporal structure of the discharge patterns reflect the amplitude envelope and periodic structure of the filtered stimulus complex (see 6, 8; 20). Animals having poorer frequency selectivity than cats and man, for example chinchillas, cannot distinguish vowels having closely spaced formants (e.g. /ae/ from /a/: 22). In impairment of hearing due to pathological conditions of the cochlea, where deterioration in frequency selectivity is predicted from the physiology (e.g. 16) and has been confirmed psychoacoustically (e.g. 30, 34, 35), impairment of intelligibility for the back vowels has been observed (e.g. 25). Here, the separation of the first two formants is likely to be too small compared with the increased peripheral filter bandwidth (see 7, and for review). Similar, but smaller, deterioration in frequency selectivity for sound complexes occurs at high sound levels (particularly 70 dB SPL and above), as measured physiologically (Figs. 1, 2 and 3 above; see 4, 19, 23) and psychophysically (29).

References

1. J.T. ALLANSON and I.C. WHITFIELD 1956 in: C. Cherry (Ed.), 3rd London Symposium on Information Theory, Butterworths, London, pp. 269-284.
2. T.J.F. BUUNEN and W.S. RHODE 1978 J. Acoust. Soc. Am. 64, 772-781.
3. E.F. EVANS 1975 in: W.D. Keidel and W.D. Neff (Eds.), Handbook of Sensory Physiology, Vol. 5/2, Springer-Verlag, Heidelberg, pp. 1-108.
4. E.F. EVANS 1977a in: E.F. Evans and J.P. Wilson (Eds.), Psychophysics and Physiology of Hearing, Academic Press, London, p. 347.
5. E.F. EVANS 1977b Proc. 9th Internat. Cong. on Acoustics, Vol. of Invited Review Lectures, Spanish Acoustical Society, Madrid, pp. 55-65.
6. E.F. EVANS 1978a Audiology 17, 369-420.
7. E.F. EVANS 1978b in: C. Ludvigsen and J. Barfod (Eds.), Sensorineural Hearing Impairment and Hearing Aids, Scand. Audiol. Suppl. 6, pp. 9-44.
8. E.F. EVANS 1980a in: G. van der Brink and F. Bilsen (Eds.), International Symposium on Psychophysical, Physiological and Behavioural Studies in Hearing, Delft University Press.
9. E.F. EVANS 1982 in: Carlson, Granstrom and Lindblom (Eds.), Representation of Speech in the Peripheral Auditory System, Elsevier, Amsterdam, pp. 27-42.
10. E.F. EVANS 1982 in: Gibb and Smith (Eds.), Otolaryngology: Otology, Butterworth and Co., London, pp. 105-131.
11. E.F. EVANS 1983 in: Klinke and Hartmann (Eds.), Hearing - Physiological Bases and Psychophysics, Springer, Berlin, pp. 140-145.
12. E.F. EVANS 1980b J. Physiol. 298, 6-7P.
13. E.F. EVANS 1981 in: J. Syka and L. Aitken (Eds.), Neuronal Mechanisms of Hearing, Plenum Press, New York, pp. 69-85.
14. E.F. EVANS 1982 in: H.B. Barlow and J.D. Mollon (Eds.), The Senses, Cambridge University Press, Cambridge, pp. 307-331.
15. E.F. EVANS and A.R. PALMER 1975 J. Physiol. 252, 60-62P.
16. E.F. EVANS and J.P. WILSON 1973 in: A.R. Moller (Ed.), Basic Mechanisms in Hearing, Academic Press, New York, pp. 519-551.
17. J.M. GOLDBERG and P.B. BROWN 1969 J. Neurophysiol. 32, 613-636.
18. J.L. GOLDSTEIN and N.Y.S. KIANG 1968 Proc. IEEE 56, 981-992.
19. R.V. HARRISON and E.F. EVANS 1982 Hearing Research 6, 303-314.
20. E. JAVEL 1980 J. Acoust. Soc. Am. 68, 133-146.

21. D. JOHNSON-DAVIES and R.D. PATTERSON 1979 J. Acoust. Soc. Am. 65, 765-770.
22. J.D. MILLER 1977 in: T.H. Bullock (Ed.), Recognition of Complex Acoustic Signals, Dahlem Konferenzen, Berlin.
23. A.R. MOLLER 1977 J. Acoust. Soc. Am. 62, 135-142.
24. B.C.J. MOORE 1978 J. Acoust. Soc. Am. 63, 524-532.
25. E. OWENS, C.B. TALBOTT and E.D. SCHUBERT 1968 J. Speech and Hearing Res. 11, 648-655.
26. A.R. PALMER and E.F. EVANS 1982 Hearing Research 6, 303-314.
27. J.O. PICKLES 1975 Acta Otolaryngol. 80, 245-254.
28. J.O. PICKLES 1979 J. Acoust. Soc. Am. 66, 1725-1732,
29. G.F. PICK 1980 J. Acoust. Soc. Am. 68, 1085-1095.
30. G.F. PICK, E.F. EVANS and J.P. WILSON 1977 in: E.F. Evans and J.P. Wilson (Eds.) Psychophysics and Physiology of Hearing, Academic Press, London, pp. 273-281.
31. J.E. ROSE, L.M. KITZES, M.M. GIBSON and J.E. HIND 1974 J. Neurophysiol. 34, 685-699.
32. M.B. SACHS and N.Y.S. KIANG 1968 J. Acoust. Soc. Am. 43, 1120-1128.
33. M.B. SACHS and E.D. YOUNG 1979 J. Acoust. Soc. Am. 66, 470-479.
34. K. SCHORN, H. WURZER, M. ZOLLNER and E. ZWICKER 1977 Lar. Rhinol. Otol. 56, 121-127.
35. F. WIGHTMAN, T. MCGEE and M. KRAMER 1977 in: E.F. Evans and J.P. Wilson (Ed (Eds.), Psychophysics and Physiology of Hearing, Academic Press, London, pp. 295-306.
36. E.D. YOUNG and M.B. SACHS 1979 J. Acoust. Soc. Am. 66, 1381-1403.
37. E. ZWICKER 1974 in: E. Zwicker and E. Terhardt (Eds.) Facts and Models Hearing, Springer-Verlag, Heidelberg, pp. 132-140.