PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects

Stuart Rosen

Phil. Trans. R. Soc. Lond. B 1992 **336**, 367-373
doi: 10.1098/rstb.1992.0070

Temporal information in speech: acoustic, auditory and linguistic aspects

STUART ROSEN

Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, U.K.

SUMMARY

The temporal properties of speech appear to play a more important role in linguistic contrasts than has hitherto been appreciated. Therefore, a new framework for describing the acoustic structure of speech based purely on temporal aspects has been developed. From this point of view, speech can be said to be comprised of three main temporal features, based on dominant fluctuation rates: envelope, periodicity, and fine-structure. Each feature has distinct acoustic manifestations, auditory and perceptual correlates, and roles in linguistic contrasts. The applicability of this three-featured temporal system is discussed in relation to hearing-impaired and normal listeners.

1. INTRODUCTION

One of the most important properties of the normal auditory mechanism is that it acts as a frequency analyser. Therefore, when exploring the relationship between the perceptual attributes of speech sounds and their acoustic structure, most emphasis is placed on the frequency spectrum. So, for example, we talk about the frequencies of the formants in a vowel, or the multi-harmonic nature of voiced speech. Recently, however, there has been much greater interest in the purely temporal properties of speech sounds, for three main reasons.

First, from psychoacoustical studies, there is now a general consensus that place–frequency mechanisms on their own can not account for many aspects of the perception of pitch, and by implication the perception of intonation in speech (for relevant reviews, see Moore & Glasberg (1986); Rosen & Fourcin (1986)). Further confirmation of the importance of temporal features in pitch perception comes from Assmann & Summerfield's (1990) attempts to model the perception of concurrent vowels which have different fundamental frequencies. They found that only models which used time, as well as place, information, could account even reasonably well for the performance of human listeners. There is also good evidence from studies on the perception of phase manipulations that temporal factors play a role even in the perception of timbre, an attribute most commonly associated with the spectral shape of a sound (see Darwin & Gardner 1986; Patterson 1987; Rosen 1986, 1987).

Second, theoretical models derived from physiological evidence suggest that temporal information is important both for the perception of melodic pitch and for the auditory representation of spectral shape (Sachs & Miller 1985; Sachs *et al.* 1983).

Finally, and perhaps most strikingly, there is the evident success of the large number of patients who have received single-channel cochlear implants. Such systems deliver an electrical signal based on the speech waveform to a single electrode placed in or near the cochlea, thus allowing no place-based frequency analysis. Yet, many of these patients have performed surprisingly well, even to the extent of being able to understand unknown sentences on the basis of an auditory signal alone (Hochmair-Desoyer *et al.* 1980, 1985). Furthermore, it appears that a number of implanted children are able to perform significantly better than any user implanted as an adult, presumably due to the greater neuronal plasticity of children (Luxford *et al.* 1987). These results have stimulated hypotheses about the extent to which purely temporal information can be effective in speech perception, and led to experiments in which normally hearing listeners are asked to identify speech sounds processed so as to contain temporal variations, but not spectral ones (Van Tasell *et al.* 1987; Rosen 1989). There has, however, been little detailed consideration of the temporal structure of speech (as displayed directly in the waveforms of figure 1) and its relation to various linguistic contrasts.

2. A FRAMEWORK FOR DESCRIBING TEMPORAL INFORMATION IN SPEECH

Therefore, as a complement to the standard Fourier-based spectral approach, we have developed a three-way partition of the temporal structure of speech based on dominant temporal fluctuation rates. Without implying anything about the extent to which this system is generally applicable (in particular to normal listeners), we now describe the three features – envelope, periodicity and fine-structure – in relation to their acoustic, auditory and perceptual and linguistic correlates. In a later section, we discuss the appro-

Phil. Trans. R. Soc. Lond. B (1992) **336,** 367–373
Printed in Great Britain

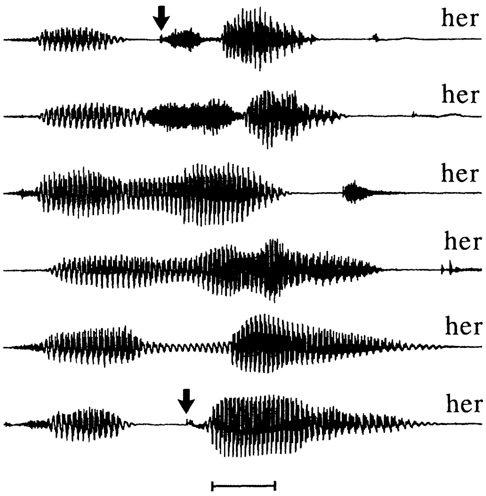

Figure 1. Speech pressure waveforms of six phrases uttered by a male adult speaker. As postvocalic 'r' is not pronounced in the speech of this speaker (as is generally the case for so-called Received Pronunciation), the initial consonants of the second word in each phrase are in an inter-vocalic position. For later reference, the two arrows (in 'chop' and 'pool') indicate the release bursts typical of plosive type sounds.

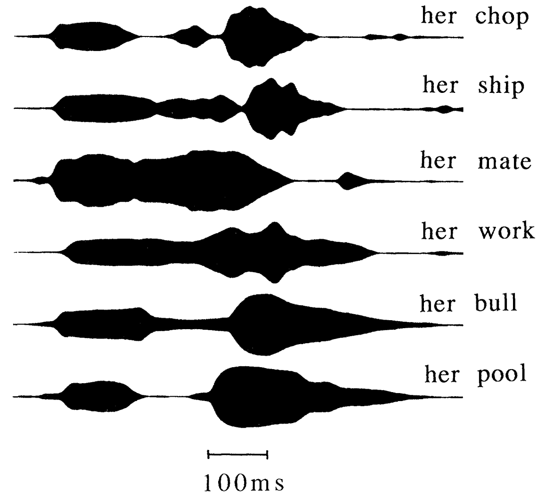

Figure 2. Waveforms obtained from those of figure 1 by full-wave rectification and low-pass filtering at 20 Hz. This preserves much of the envelope information, but eliminates that occurring at relatively high fluctuation rates. Note, for example, the loss of the release bursts of /tʃ/ in 'chop' and /p/ in 'pool' evident in figure 1.

priateness of using a temporal feature system in describing the speech-perceptual performance of various categories of hearing-impaired and normal listener.

(*a*) *Envelope*

We will refer to fluctuations in overall amplitude at rates between about 2 and 50 Hz as envelope information†. Variously known as 'amplitude envelope', 'time–amplitude', or 'time-intensity' information, this is typically what is meant by 'temporal information' in much of the literature. Envelope may be described mainly by such acoustic features as intensity, duration, rise time and fall time. Its main auditory correlates are loudness, length, attack and decay. Envelope's low frequency variations can convey four main types of linguistic information.

(*i*) *Segmental cues to manner of articulation*

Consider, for example, the voiceless affricate–fricative distinction (/tʃ/ and /ʃ/, contrasting the initial consonants of 'chop' and 'ship') for which a number of envelope features are known to be influential (Dorman *et al.* 1980; Gerstman 1957; Howell & Rosen 1983, 1987; Repp *et al.* 1978). The frication noise of /tʃ/ has a quicker rise time and shorter overall duration than the corresponding frication of /ʃ/. /tʃ/ has a short release burst whereas /ʃ/ does not. Short release bursts (transients) typically indicate plosive type sounds. Silent gaps too may indicate the presence of a voiceless plosive or affricate (as in the intervocalic /tʃ/ and /p/ of figure 2: Bailey & Summerfield 1980; Summerfield *et al.* 1981). More generally, it has been proposed that relatively fast changes in overall amplitude mark consonants from non-consonants (Stevens 1980, 1981; Stevens & Blumstein 1981), or continuants from non-continuants (Shinn & Blumstein 1984).

(*ii*) *Segmental cues to voicing*

Voiced sonorants (vowels, semivowels, nasals and laterals) typically have greater amplitudes than voiceless obstruents (as in the /m/ of 'mate' versus the /p/ of 'pool' in figure 2). The existence and duration of silent intervals may be important in distinguishing voiced from voiceless plosives in intervocalic position (as in the figured /b/ of 'bull' versus /p/ of 'pool'). In some contexts, vowel duration, in so far as it is cued by envelope, can give information about voicing in the following consonant (Umeda 1975). Generally speaking, however, the envelope cues to voicing contrasts are weak.

(*iii*) *Segmental cues to vowel identity*

The duration of vowels varies lawfully with vowel quality, and so can signal some information about it, albeit weakly. Many languages use duration contrastively (along with changes in quality) to distinguish among vowels (see Lehiste (1970, pp. 18–19, 30–35) for a review). For example, other things being equal, the vowel in'heed' tends to be of significantly longer duration than that of 'hid'.

(*iv*) *Prosodic cues*

Dynamic envelope cues can be used to assist syllabification (as Mermelstein (1975) has shown in an automated procedure), and relative amplitude (on a more static basis) probably plays a minor role in the assignment of stress in words (for example in dis-

† Note that this 'envelope' is distinct from, although related to, the 'envelope' derived through use of the analytic signal. In the terms used here, the 'envelope' derived from the analytic signal typically contains both envelope and periodicity information (see Seggie 1986*a*, *b*).

tinguishing the verb 'rebel' from the noun 'rebel': see Crystal (1969, pp. 113–120); Fry (1968): Lehiste (1970, pp. 36–38, 120–139) for reviews of the relevant literature). In so far as amplitude onsets and offsets can demarcate linguistic units (vowel, syllable or word), much information about duration, and hence speech rhythm and tempo can also be extracted from envelope cues. Duration itself appears to play a role in word-level stress (for example, as in 'rebel' versus 'rebel') whereas information about tempo could assist listeners in normalizing for speech rate variations in segmental (Miller 1981) and prosodic contrasts. Variations in speech rate can also carry distinctions in meaning (Crystal 1969, pp. 152–156) or indicate parenthetical comments.

(*b*) *Periodicity*

Properties of the speech signal which relate to the distinction between periodic and aperiodic stimulation, and to the rate of periodic stimulation, shall be referred to simply as periodicity information. Periodic sounds fluctuate primarily at rates between about 50 and 500 Hz, whereas aperiodic sounds typically fluctuate at rates from a few kHz up to 5–10 kHz (although they can fluctuate at rates below 1 kHz). Because periodic and aperiodic sounds generally differ so greatly in their characteristic rates of fluctuation, it may be useful at times to think of periodicity information as being divided into two subclasses (periodic and aperiodic) with different dominant fluctuation rates, but which both give information about the source of excitation in speech production. Note, however, that the ability to distinguish periodic from aperiodic stimulation does not necessarily depend upon differential sensitivity to high fluctuation rates (that is, the ability to distinguish among sounds which fluctuate at different high rates). It is enough to distinguish absence of stimulation (indicating silence) from stimulation with low-frequency periodicity (below, say, 1 kHz, indicating a periodic sound) from stimulation of a high (but otherwise indeterminate) fluctuation rate (indicating an aperiodic sound). Similarly, detection of aperiodicity may rely on the irregular nature of the low-frequency fluctuations in a rapidly fluctuating sound.

The acoustic contrast of periodicity versus aperiodicity is reflected in the time domain as regularity versus irregularity of the speech signal, and in the frequency domain as the distinction between a harmonic and a continuous spectrum. Auditorily, this contrast is perceived as one of buzziness versus noisiness. Changes in the rate of periodic fluctuations are reflected in changes of fundamental frequency, which lead to a sensation of melodic pitch. Periodicity information directly conveys two main types of linguistic information.

(*i*) *Segmental information about voicing and manner*

The presence of low-frequency quasi-periodic acoustic energy in a speech signal is a reflection of the quasi-periodic vibrations of the vocal folds (as in the /m/ of mate in figure 3). Such sounds are said to be voiced, and such voicing is the most important cue to the phonological feature of voicing, perhaps the most basic distinction in all of the world's languages. Similarly, in many languages (such as English and French) there is an association between manner and voicing features (e.g. all nasals are voiced) which permits manner information to be obtained from information about phonetic voicing patterns. Speech segments which are aperiodic result from turbulence noise generated by aerodynamic flow between closely spaced articulators. Such aperiodicity can be a strong cue for voicelessness, or to the fricative manner of articulation (as in the /ʃ/ of 'ship' in figure 3).

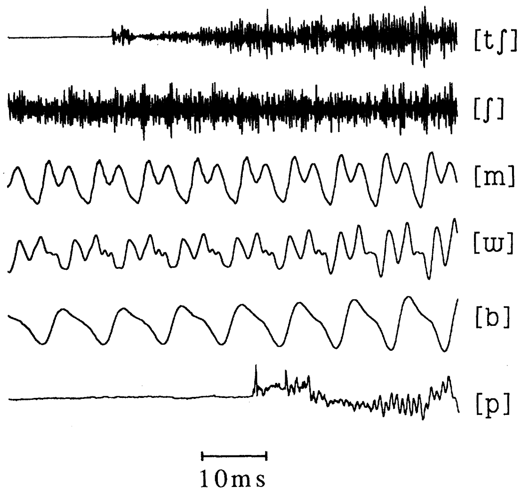

Figure 3. Sections of the speech pressure waveforms of figure 1, chosen so as to illustrate periodicity information. Note the irregular waveforms associated with the voiceless sounds /tʃ/, /ʃ/ and /p/, and the regularity exhibited by the quasi-periodic voiced sounds /m/, /w/ and /b/.

(*ii*) *Prosodic information relating to intonation and stress*

The fundamental frequency of quasi-periodic energy in a speech signal is a reflection of the rate of vocal fold vibration, and is the prime acoustic correlate of the perception of voice pitch. Linguistically meaningful patterns of voice pitch are known as intonation and tone, and play important roles in accenting syllables in words and sentences, in clarifying ambiguous pronoun references, in marking syntactic units and in distinguishing questions and statements (for reviews see Fry (1968); Lehiste (1970); Rosen & Fourcin (1986)). Furthermore, in tone languages like Chinese, voice pitch patterns have a lexical function: that is, they distinguish different dictionary meanings of a word. For example, in Cantonese Chinese, the syllable 'yee' may mean 'clothes', 'chair', 'meaning', 'child', 'ear' or 'two', depending upon the pitch contour used when uttering it. Even English has a minor instance of this, in that voice pitch contours can play an important role in distinguishing between the verbal and nominal function of a word (as in 'rebel' versus 'rebel': see above).

(*c*) *Fine-structure*

We shall refer to variations of wave shape within

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

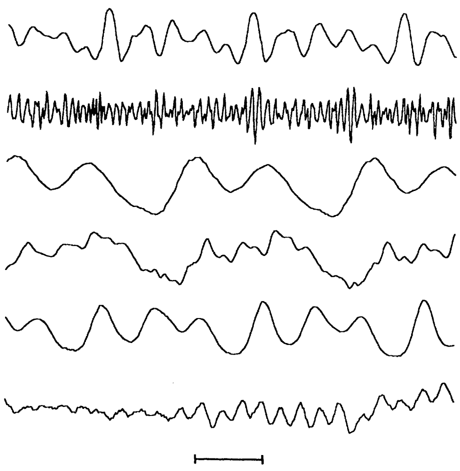

Figure 4. Sections of the speech pressure waveforms of figure 1, chosen so as to illustrate fine-structure information. Note the differences in wave shape among the four vowel-like sounds, and the generally higher fluctuation rate of /ʃ/ as compared to /p/, a reflection of the fact that /p/ has its spectral energy concentrated at lower frequencies than does /ʃ/.

single periods of periodic sounds, or over short time intervals of aperiodic ones as fine-structure information. This cue has dominant fluctuation rates from about 600 Hz to about 10 kHz. Acoustically, fine-structure informs about the spectrum of a sound (both amplitude and phase)‡ and so also contains its formant pattern. Fine-structure's primary auditory correlate relates to timbre, or quality, and it can convey at least two types of segmental linguistic information.

(i) Segmental cues to place of articulation and vowel quality

This potential function of fine-structure is by far the most important, not least because spectral shape variations are more or less the only acoustic cues to place. For example, it is well known that the two most important acoustic features that distinguish the words 'bait', 'date' and 'gate' from one another are the frequency spectrum of the initial release burst, and the dynamic formant transitions that follow (see, for example, Hazan & Rosen (1991) and references therein), all contained in fine-structure cues. Similarly, English voiceless fricatives in a prevocalic position may be distinguished from one another on the basis of static spectral shape, or the formant transitions in the following vowel, with the importance of each cue strongly dependent on the particular place of articulation (Harris 1958). Finally, spectral shape is the major cue to vowel identity (for examples, see figure 4).

‡ As no linguistic contrast depends primarily upon changes in phase spectrum, we discuss it no longer. It is important to realize, however, that the temporal structure of a speech signal (and especially its fine structure) will change with changes in the phase spectrum. So, the more a listener is relying on temporal cues for perception, the more important phase will be in determining auditory percepts (for a related point, see Rosen & Fourcin (1986)).

LINGUISTIC CONTRAST		TEMPORAL FEATURES: envelope	periodicity	fine-structure
segmental	manner	★	★	⋆
	voicing	⋆	★	★
	place			★
	voice quality	⋆		★
prosodic	tempo, rhythm	★		
	syllabicity	★		
	stress	⋆	★	
	intonation		★	

Table 1. The potential role of various temporal features of speech in linguistic contrasts. The size of the stars indicates the extent to which a particular feature operates in a particular linguistic contrast, with a blank space indicating very weak or non-existent cues. Note that prosodic cues, which by definition occur across a number of segments, tend to be cued by the lower fluctuation rate categories. In fact, even for a linguistic feature such as intonation, related to periodicity, the relevant patterning (the fundamental frequency contour) occurs over much slower timescales than periodicity itself.

(ii) Segmental cues to voicing and manner

Voiced sounds have a spectrum heavily weighted to low frequencies (below about 1 kHz), and hence tend to have low fluctuation rates, whereas voiceless sounds typically have their peak energies at considerably higher frequencies, and thus tend to have high fluctuation rates (see, for example, the differences between the vowels and /ʃ/ in figure 4). First-formant transitions are known to play some role in distinguishing English voiced from voiceless plosives in initial prevocalic position (e.g. Soli 1983; Stevens & Klatt 1974). Apart from the manner information signalled by voicing, other cues to manner may be signalled by the shape of the spectrum. Nasals, for example, are characterized by a low first formant frequency, broad resonances, and zeros in the spectrum (Fujimura 1962). Stevens (1980, 1981) has discussed the role of sudden spectral changes (usually in conjunction with sudden envelope changes) in distinguishing consonantal sounds from non-consonantal ones.

3. CAVEATS

This list of correspondences between temporal features and linguistic information is not exhaustive, as there are weak potential cues to speech contrasts that are not mentioned above §. However, the most important

§ For example, fine structure may contain weak prosodic cues. At least in English, the vowels in unstressed syllables tend to be more neutral in quality than the vowels in stressed syllables. There is some evidence that this feature influences decisions of the 're<u>bel</u>' versus '<u>re</u>bel' variety (see above). So, too, may envelope give weak information about place of articulation, because the duration of aperiodicity in voiceless initial English plosives is known to vary lawfully with place of articulation (Lisker & Abramson 1964).

relationships are summarized in table 1. Note that temporal features may not always operate independently of one another. Duration, which has been grouped as an envelope cue, usually refers to the duration of an interval of speech with particular acoustic properties. For example, it is the duration of aperiodicity that helps to distinguish /tʃ/ from /ʃ/. Finally, there is likely to be much overlap in the frequency region over which the features operate. The release burst which is present in /tʃ/ but not in /ʃ/ is so short that its envelope certainly contains frequencies above 50 Hz. It is still the case, however, that the properties of the burst and affrication can be reasonably well divided among the features of envelope, periodicity and fine-structure.

4. APPLYING THE FRAMEWORK TO VARIOUS KINDS OF LISTENERS

When we come to consider the role of temporal features in speech perception, complications arise depending upon the type of listeners we are dealing with. For users of single-channel implants, no place–frequency mechanisms operate and so all auditory perception must be based on temporal features. Here, the three-way framework detailed above will apply completely, and the only extra problem concerns the extent to which temporal features are modified by the patient's speech processor. Leaving this difficulty aside, it appears that most users of analogue single-channel implants can use envelope and periodicity information in a linguistic way, but the relative importance of each cue is not yet clear (Agelfors & Risberg 1989; Rosen & Ball 1986; Rosen *et al.* 1989; Tyler *et al.* 1987). In many, if not most, patients, there is also sensitivity to temporal fine-structure, but this is relatively rarely used in the perception of natural speech (Agelfors & Risberg 1987, 1989; Hochmair-Desoyer *et al.* 1985; Rosen & Ball 1986; Rosen *et al.* 1989; White 1983). However, the reception of unknown sentences (Hochmair-Desoyer *et al.* 1980, 1985) or accurate identification of place of articulation in consonants (Shannon *et al.* 1992) by auditory means alone implies some linguistic use of the temporal fine structure of speech.

At the other end of the observer continuum lie normal listeners, in whom the effects of peripheral auditory filtering must be considered||. The normal auditory system decomposes a speech waveform, via the filtering action of the cochlea, into many waveforms, each of which will have its own three-way complement of temporal information (see figure 5). So, for instance, the envelope features transmitted by the auditory nerve will be modifications (to a greater

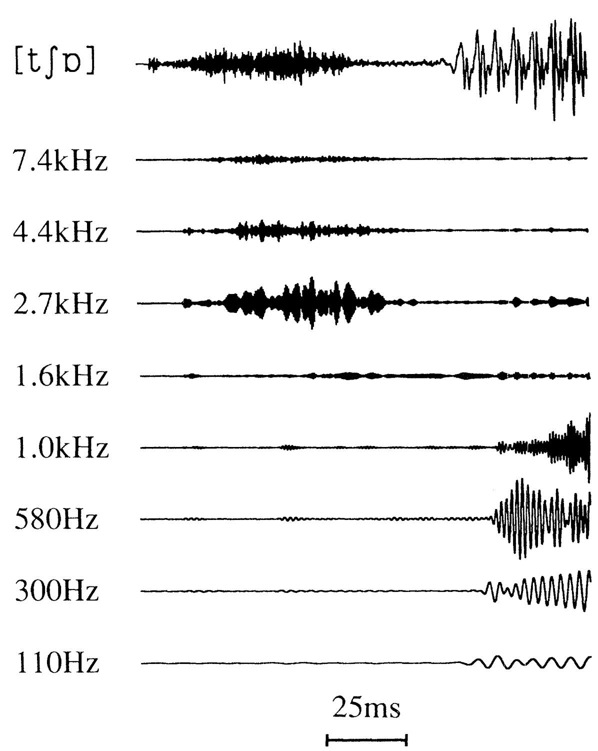

Figure 5. Original speech waveform of the initial part of the word 'chop', and waveforms resulting from auditory-like linear filtering at various centre frequencies. Note that the envelope features relating to the release burst, rise time and duration of the frication can only be found at the higher centre frequencies. At the same time, the periodicity of the vowel is only evident at lower centre frequencies.

|| Plomp (1983) has described a three-way partition of the properties of speech sounds based on the concept of modulation which, although it does not purport to be a purely temporally-based description of speech, bears some resemblance to the system described here. In Plomp's words: 'Speech can be considered to be a wide-band complex signal modulated continuously in time in three different respects: (i) the vibration frequency of the vocal cords is modulated, determining the pitch variations of the voice; (ii) the temporal envelope of this signal is modulated by narrowing and widening the vocal tract locally by means of the tongue and lips; and (iii) the tongue and the lips in combination with the cavities of the vocal tract determine the sound spectrum of the speech signal, which may be considered as a modulation along the frequency scale.' These are clearly related to, although not identical with, the periodicity, envelope and fine-structure categories described above. There are, however, a number of difficulties in Plomp's characterization.

1. Fine structure is only discussed via the frequency domain (i.e. as a spectrum). Although perhaps reasonable for a system aimed solely at explaining the perception of normal listeners, it limits its usefulness as regards cochlear implants and theories of normal hearing which require use of temporal features in the perception of spectral shape variations.

2. No mechanism for distinguishing periodic from aperiodic sounds is proposed (the latter are never mentioned), nor is there any appropriate 'dimension' (i.e. frequency or time) suggested for the representation of periodicity (apart from its modulation in period over time). In fact, there appears to be the implication that all speech sounds are voiced.

3. Aspects of envelope are controlled by vocal fold behaviour, and not only by supralaryngeal manoeuvres (e.g. the decrease in amplitude over the vowel for any CV syllable uttered in citation form, as in 'key')

4. The spectrum of voiced speech sounds is influenced by the spectrum of the source (determined by vocal fold behaviour), as well as by the shape of the vocal tract. Generally speaking, it appears that a description in terms of acoustic properties is more useful than a description in terms of production, not least because it eliminates descriptive difficulties like the last two just mentioned. In terms of traditional descriptions of speech production, only the temporal feature of periodicity has a simple productive correlate: that of the source of excitation. Fine structure results, as implied above, from the interaction between source and filter, as indeed does envelope.

or lesser extent) of those observed on a speech waveform.

More importantly, peripheral auditory filtering means that temporal cues will be transformed into place cues, at least for periodicity and fine-structure. Although there is strong evidence that temporal properties of the signal have some role in the perception of periodicity and fine structure even in the normal listener, no one doubts that peripheral place–frequency analysis is crucial. Only for envelope features does it appear that temporal processes are nearly completely dominant¶.

Somewhere in the middle of this continuum on which listeners vary in the extent of their place–frequency analysis are hearing-impaired listeners and users of multi-channel implants. Moderately impaired listeners can retain a high degree of frequency selectivity, while those with more profound impairments can exhibit little or none (Faulkner *et al.* 1990). Multi-channel implant users, whose frequency selectivity is typically based on electronic filters in their speech processors, are probably comparable in selectivity to listeners who are severely or profoundly hearing impaired†. In any case, the greater the degradation of auditory frequency selectivity, the greater the role of explicitly temporal factors.

There is also a general limitation for any auditory system in the use of high fluctuation rates due to neurophysiological constraints on temporal coding. In particular, it is well known that mammalian auditory nerve fibres only synchronize their firing to acoustic sinusoids up to a maximum of about 5 kHz, so this must be an absolute upper limit on differential sensitivity to fine-structure and periodicity fluctuations (a limit that is probably also applicable to electrical stimulation).

5. FINAL REMARKS

There is still much to be learned about the role of the temporal structure of speech in influencing speech perception. Further clarification will lead not only to theoretical advances in understanding normal hearing, but also allow more efficient utilisation of limited auditory capacities in users of both acoustic and electro-cochlear auditory prostheses.

Thanks to Richard Baker, who assisted in making the figures, and Beverly Wright, Ken Robinson, Mike Johnson, Andy Faulkner, Bill Barry, Ginny Ball, Richard Baker, Michael Ashby and Evelyn Abberton for useful discussions or suggestions which greatly improved this paper. This work is supported by the Medical Research Council (U.K.).

¶ Even the auditory representation of envelope can be affected in minor ways by place–frequency analysis. For example, an aperiodic sound which is part of a burst of short duration will have a wider spectrum than the same sound turned on gradually.

† One important difference between implant users and profoundly hearing-impaired listeners is that the latter typically have a much more restricted range over which auditory stimulation is possible (perhaps only up to 1 kHz or so; Rosen *et al.* 1987). This kind of 'low-pass filtering' may have important implications for the representation of various temporal features (e.g. in periodicity versus aperiodicity, as most aperiodic energy is above 1 kHz).

REFERENCES

Agelfors, E. & Risberg, A. 1987 The identification of synthetic vowels by patients using a single-channel cochlear implant. *Proceedings of the XIth International Congress of Phonetic Sciences* **4**, pp. 181–184, Tallinn. (Also published in Speech Transmission Laboratory: Quarterly Progress and Status Report **2–3**, 31–38. Stockholm: Royal Institute of Technology. (1987))

Agelfors, E. & Risberg, A. 1989 Speech feature perception by patients using a single-channel Vienna 3M extracochlear implant. *Proceedings of the Speech Research '89 International Conference*, pp. 149–152. Budapest: Linguistics Institute of the Hungarian Academy of Sciences. (Also published in Speech Transmission Laboratory: Quarterly Progress and Status Report **1**, 145–149. Stockholm: Royal Institute of Technology. (1989))

Assmann, P.F. & Summerfield, Q. 1990 Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J. acoust. Soc. Am.* **88**, 680–697.

Bacon, S.P. & Viemeister, N.F. 1985 Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology* **24**, 117–134.

Bailey, P.J. & Summerfield, Q. 1980 Information in speech: observations on the perception of [s]-stop clusters. *J. exp. Psychol.: Hum. Percept. Perform.* **6**, 536–563.

Burns, E.M. & Viemeister, N.F. 1976 Nonspectral pitch. *J. acoust. Soc. Am.* **60**, 863–869.

Crystal, D. 1969 *Prosodic systems and intonation in English.* Cambridge University Press.

Darwin, C.J. & Gardner, R.B. 1986 Mistuning a harmonic of a vowel: grouping and phase effects on vowel quality. *J. acoust. Soc. Am.* **79**, 838–845.

Dorman, M.F., Raphael, L.J. & Isenberg, D. 1980 Acoustic cues for a fricative-affricate contrast in word-final position. *J. Phonet.* **8**, 397–405.

Faulkner, A., Rosen, S. & Moore, B.C.J. 1990 Residual frequency selectivity in the profoundly hearing-impaired listener. *Br. J. Audiol.* **24**, 381–392.

Fry, D.B. 1968 Prosodic phenomena. In *Manual of phonetics* (ed. B. Malmberg), pp. 365–410. Amsterdam: North Holland.

Fujimura, O. 1962 Analysis of nasal consonants. *J. acoust. Soc. Am.* **34**, 1865–1875. (Also published in *Readings in acoustic phonetics* (I. Lehiste), pp. 238–248. Cambridge, Massachusetts: MIT Press. (1967))

Gerstman, L.J. 1957 Perceptual dimensions for the friction portions of certain speech sounds. Ph. D. thesis, New York University.

Harris, K.S. 1958 Cues for the discrimination of American English fricatives in spoken syllables. *Lang. Speech* **1**, 1–7. (Also published in *Acoustic phonetics* (ed. D. B. Fry), pp. 284–297. Cambridge University Press. (1976))

Hazan, V. & Rosen, S. 1991 Individual variability in the perception of cues to place contrasts in initial stops. *Percept. Psychophys.* **49**, 187–200.

Hochmair-Desoyer, I.J., Hochmair, E.S., Fischer, R.E. & Burian, K. 1980 Cochlear prostheses in use: recent speech comprehension results. *Arch. Oto-Rhino-Laryngol.* **229**, 81–98.

Hochmair-Desoyer, I.J., Hochmair, E.S. & Stiglbrunner, H.K. 1985 Psychoacoustic temporal processing and speech understanding in cochlear implant patients. In *Cochlear implants* (ed. R. A. Schindler & M. M. Merzenich), pp. 291–304. New York: Raven Press.

Howard, D.M. & Seligman, P.M. 1983 Initial comparisons between two simple time-domain fundamental frequency detectors. *Speech Hear. Lang.* (Prog. Rep. Dept. Phonetics & Linguistics, Univ. Coll. Lond.) **1**, 95–105.

Howell, P. & Rosen, S. 1983 Production and perception of rise time in the voiceless affricate/fricative distinction. *J. acoust. Soc. Am.* **73**, 976–984.

Howell, P. & Rosen, S. 1987 Perceptual integration of rise time and silence in affricate/fricative and pluck/bow continua. In *The psychophysics of speech perception* (ed. M. E. H. Schouten), pp. 173–180. Dordrecht: Martinus Nijhoff.

Lehiste, I. 1970 *Suprasegmentals*. Cambridge, Massachusetts: MIT Press.

Lisker, L. & Abramson, A.S. 1964 A cross-language study of voicing in initial stops: Acoustical measurements. *Word* **20**, 384–422.

Luxford, W.M., Berliner, K.I., Eisenberg, L.S. & House, W.F. 1987 Cochlear implants in children. *Annls Otol. Rhinol. Laryngol.* **96**(Suppl. 128), 136–138.

Mermelstein, P. 1975 Automatic segmentation of speech into syllabic units. *J. acoust. Soc. Am.* **58**, 880–883.

Miller, J.L. 1981 Effects of speaking rate on segmental distinctions. In *Perspectives on the study of speech* (ed. P. D. Eimas & J. L. Miller), pp. 39–74. Hillsdale, New Jersey: Lawrence Erlbaum.

Moore, B.C.J. & Glasberg, B.R. 1986 The role of frequency selectivity in the perception of loudness, pitch and time. In *Frequency selectivity in hearing* (ed. B. C. J. Moore), pp. 251–308. London: Academic Press.

Patterson, R.D. 1987 A pulse ribbon model of monaural phase perception. *J. acoust. Soc. Am.* **82**, 1560–1586.

Plomp, R. 1983 The role of modulation in hearing. In *Hearing – physiological bases and psychophysics* (ed. R. Klinke & R. Hartmann), pp. 270–276. Berlin: Springer-Verlag.

Repp, B.H., Liberman, A.M., Eccardt, T. & Pesetsky, D. 1978 Perceptual integration of acoustic cues for stop, fricative and affricate manner. *J. exp. Psychol.: Hum. Percept. Perform.* **4**, 621–637.

Rosen, S. 1986 Monaural phase sensitivity: frequency selectivity and temporal processes. In *Auditory frequency selectivity* (ed. B. C. J. Moore & R. D. Patterson), pp. 419–426. New York: Plenum.

Rosen, S. 1987 Phase and the hearing-impaired. In *The psychophysics of speech perception* (ed. M. E. H. Schouten), pp. 481–488. Dordrecht: Martinus Nijhoff.

Rosen, S. 1989 Temporal information in speech and its relevance for cochlear implants. In *Cochlear implant: acquisitions and controversies* (ed. B. Fraysse & N. Cochard), pp. 3–26. Basel: Cochlear AG.

Rosen, S. & Fourcin, A.J. 1986 Frequency selectivity and the perception of speech. In *Frequency selectivity in hearing* (ed. B. C. J. Moore), pp. 373–487. London: Academic Press.

Rosen, S. & Ball, V. 1986 Speech perception with the Vienna extra-cochlear single-channel implant: a comparison of two approaches to speech coding. *Br. J. Audiol.* **20**, 61–83.

Rosen, S., Walliker, J.R., Brimacombe, J.A. & Edgerton, B.E. 1989 Prosodic and segmental aspects of speech perception with the House/3M single-channel implant. *J. Speech Hear. Res.* **32**, 93–111.

Rosen, S., Walliker, J.R., Fourcin, A.J. & Ball, V. 1987 A microprocessor-based acoustic hearing aid for the profoundly impaired listener. *J. Rehab. Res. Develop.* **24**, 239–260.

Sachs, M.B. & Miller, M.I. 1985 Pitch coding in the auditory nerve: Possible mechanisms of pitch sensation with cochlear implants. In *Cochlear implants* (ed. R. A. Schindler & M. M. Merzenich), pp. 185–194. New York: Raven Press.

Sachs, M.B., Young, E.D. & Miller, M.I. 1983 Speech encoding in the auditory nerve: Implications for cochlear implants. In *Cochlear prostheses, an international symposium* (ed. by C. W. Parkins & S. W. Anderson) (*Ann. N.Y. Acad. Sci.* **405**), pp. 94–114. New York Academy of Sciences.

Seggie, D. 1986*a* The use of signal instantaneous frequency for voicing determination. *Speech Hear. Lang.* (Prog. Rep. Dept. Phonetics & Linguistics, Univ. Coll. Lond.) **2**, 179–192.

Seggie, D. 1986*b* The application of analytic signal analysis in speech processing. *Proceedings of the Institute of Acoustics Autumn Conference: Speech and Hearing, Windermere, U.K.* **8**, 85–92.

Shannon, R.V., Zeng, F.-G. & Wygonski, J. 1992 Speech recognition using only temporal cues. In *The auditory processing of speech: from sounds to words* (ed. M. E. H. Schouten). Berlin: Mouton De Gruyter. (In the press.)

Shinn, P. & Blumstein, S.E. 1984 On the role of the amplitude envelope for the perception of [b] and [w]. *J. acoust. Soc. Am.* **75**, 1243–1252.

Soli, S.D. 1983 The role of spectral cues in discrimination of voice onset time differences. *J. acoust. Soc. Am.* **73**, 2150–2165.

Stevens, K.N. 1980 Acoustic correlates of some phonetic categories. *J. acoust. Soc. Am.* **68**, 836–842.

Stevens, K.N. 1981 Constraints imposed by the auditory system on the properties used to classify speech sounds: data from phonology, acoustics and psycho-acoustics. In *The cognitive representation of speech* (ed. T. F. Myers, J. Laver & J. Anderson), pp. 61–74. Amsterdam: North Holland.

Stevens, K.N. & Blumstein, S.E. 1981 The search for invariant acoustic correlates of phonetic features. In *Perspectives on the study of speech* (ed. P. D. Eimas & J. L. Miller), pp. 1–38. Hillsdale, New Jersey: Lawrence Erlbaum.

Stevens, K.N. & Klatt, D.H. 1974 Role of formant transitions in the voiced-voiceless distinction for stops. *J. acoust. Soc. Am.* **55**, 653–659.

Summerfield, Q., Bailey, P.J., Seton, J. & Dorman, M.F. 1981 Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split'. *Phonetica* **38**, 181–192.

Tyler, R.S., Tye-Murray, N., Preece, J.P., Gantz, B.J. & McCabe, B.F. 1987 Vowel and consonant confusions among cochlear implant patients: do different implants make a difference? *Annls Otol. Rhinol. Laryngol.* **96**(Suppl. 128), 141–144.

Umeda, N. 1975 Vowel duration in American English *J. acoust. Soc. Am.* **58**, 434–445.

Van Tasell, D.J., Soli, S.D., Kirby, V.M. & Widin, G.P. 1987 Speech waveform envelope cues for consonant recognition. *J. acoust. Soc. Am.* **82**, 1152–1161.

White, M.W. 1983 Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes. In *Cochlear prostheses, an international symposium* (ed. C.W. Parkins & S.W. Anderson) (*Ann. N.Y. Acad. Sci.* **405**), pp. 348–359. New York Academy of Sciences.