SYMPOSIUM 4: Auditory Analysis and Speech Perception

Chairman: *A. Fourcin, London, United Kingdom*

Panel members: *A. Risberg, J.M. Pickett and S.G. Revoile, R.J. Ritsma and J.W. Horst, J.L. Goldstein*

This symposium was divided into two clear parts. In the first, the five discussants, who had contributed invited abstracts to the main programme, each presented a summary of their individual work and gave a further overview of its more general implications.

In the second part, the general discussion touched on five main areas in which work on hearing has practical and theoretical bearing on human speech processing. Several themes were common to both aspects of the symposium. First, there was a bias towards the use of research with the hearing impaired as a means of better understanding the mechanism of normal hearing. Second, there was a general interest in models of hearing with special reference to frequency and temporal analysis and peripheral rather than cognitive processing. Third, the use of speech contrastive stimuli as an analytic - experimental - tool was discussed with particular reference to the use of recoded natural syntheses rather than complete natural speech.

It is best to follow the logical rather than the temporal organisation of these various elements.

Julius Goldstein gave a discussion of the possible mechanisms underlying our perception of voice pitch, and evaluated the broad ways in which a more complete speaker-listener model would be influenced by these basic peripheral processing issues. The essential proposal is that voice pitch is mediated by a two- fold process. First, auditory filtering corresponding to what is experimentally found neuro-physiologically in animals (Kiang, 1965) and psychophysiologically determined in man (Patterson et al., 1982), processes the complex acoustic input. Second, auditory fibres close in characteristic frequency to this region of maximum place response then give a temporal pattern of activity which is used as an input to an 'optimum' central processor of fundamental frequency (Goldstein, 1983; Goldstein and Srulovicz, 1983). Although temporal processing is, here, prominent in the last part of the peripheral sequence of operations this hypothesis does not rely on time patterning as a primary basis for sound analysis. Indeed, in the overall discussion Hoekstra's work was often referred to as demonstrating the relative unimportance of any pitch mediating algorithm which did not rely on place analysis. The relevance of situations in which no place analysis is possible and yet temporal patterning elicits useful prosodic contrastive ability (Fourcin et al., 1983) was pointed out. In normal listeners, however,

this temporal feature of auditory analysis may be of lesser importance in the developed adult. It was proposed, however, that it might be crucial both for the normally developing infant and for the congenitally hearing impaired (Fourcin, 1977).

The general discussion went from these considerations relating to pitch to deal with timbre analysis and the relevance of different auditory models. In particular the 'Stockholm' zero crossing model was compared with Goldstein's (and, effectively, Sachs and Young's 1982) model with particular reference to the possibility not only of analysing spectral prominences but also assessing gross spectral weighting.

Current work in Stockholm has led to the 'DOMIN' model (Carlson & Granström, 1982) and whilst this has the desirable feature of delineating spectral prominences the point was made in the discussion that timbre effects coming from level variations within a spectrum were possibly not sufficiently well represented to cater for all phonetic contrasts. The trend, however, from a largely physical model to one which was substantially determined by observational parameters was considered essential to future progress - in work on both man and machine.

The contributed papers by Ritsma and Horst (1983) and Pickett and Revoile (1983) related to speech processing by the normal and by the impaired auditory system. An essential implication of the auditory models discussed immediately above is that a reduction in peripheral frequency selectivity will inevitably entail a corresponding impairment of speech receptive ability. Ritsma's presentation gave experimental evidence to show, however, that essentially normal responses could be obtained even for the hearing impaired with a demonstrably degraded auditory filter bandwidth in quiet listening to a single formant changed in centre frequency with bandwidth as a control parameter. When the signal/noise ratio of the stimuli was degraded, the difference between normal and hearing impaired listeners became apparent. The immediate conclusion to be drawn from these experiments is that 'frequency selectivity primarily influences (the) speech-to-noise ratio and not (the) speech perception as such.' In the broader context of this Symposium, however, it emerges that it may be more accurate to postulate that frequency selectivity influences the complexity of speech processing available to the listener (see particularly the contributions by Risberg, 1983; Pickett et al., 1983 and by Fourcin, 1983.

Arne Risberg's contribution complemented the work presented by Ritsma and Horst in that he gave results relating another measure of frequency selectivity to speech receptive ability for a range of hearing impaired subjects. Here, scores tended to be lower as selectivity decreased even though the tests were performed without added noise. The subjects with poor selectivity all complained, however, that especially in a noisy situation they had speech perception difficulties.

Risberg also discussed the relation between prosodic receptive ability and his (frequency modulated tone) assessment of frequency. Here, a fairly clear

correlation between low frequency discrimination ability and response accuracy in stressed word identification test emerged. This finding was especially relevant to the topic of frequency versus temporal processing which was a central discussion theme and it seemed possible that, contrary to the case for normals as shown by Hoekstra (whose results were given special attention by several speakers), temporal processing might be basic to these findings (and this correlates with Fourcin's report on the EPI group's work on electro-cochlear stimulation with the totally deaf - where only temporal processing is possible - (cf. Fourcin et al., 1983)).

Measurements of this type with congenitally hearing-impaired subjects are difficult to interpret, however, since response ability is conditioned by speech knowledge as well as sensory processing.

Fourcin's contribution (1983) was especially concerned with the selective use of simple minimal pair contrasts synthesized so that only a small number of acoustic elements were involved. Using auditory tests based on the interactive use of these stimuli it was possible to follow the course of speech acquisition in both normal and hearing- impaired children and to show that for different ages, and hearing abilities, different pattern aspects of the speech signal were of importance. The tendency being for temporal and low frequency components to dominate initially and to be reinforced by higher frequency constrastive information at later stages of development. This analytic use of speech material was also of value in speech aids and examples were discussed of the use of only fundamental frequency information in both acoustic and electrical stimulation prostheses. In each case quite substantial improvement was possible not only in lip-read receptive ability but also in speaking ability.

These techniques seemed likely, in the future, to be of much wider applications since they provide ways of training and assessment which are not otherwise available and link psycho-acoustic and phonetic levels of processing.

Pickett and Revoile (1983) presented the results of experiments in which natural speech was used as the basis of tests of speech perceptual ability in both normal and hearing impaired subjects. Initial and final voiced voiceless cognate consonantal contrasts were controlled, for example, in regard to frication; appropriate onset and offset transitions; vowel duration. Fairly consistent differences in overall response strategies were found between the two groups of listeners in regard to the use of durational and low frequency speech pattern elements although some individuals showed marked differences in performance. This aspect of speech perception was emphasized also by Risberg and by Fourcin and for the latter arose from the intrinsic auditory nature of speech contrasts. When a number of potentially useful contrastive features exist - and this is always the case for normal speech - the listener must make a choice. The act of choice creates a feature sub-set which, whether synthesis modification of natural speech is used, may map quite differently from one subject to another, onto the information available in the stimulus presentation.

A final, but related, discussion topic concerned the dimensionality of the auditory representation of speech. One argument proposed the use of factor analysis techniques (cf. Pols and Stoop, 1983) to arrive at a pertinent set of contrastive features which would be relevant to the set of listeners investigated. Another proposition was that the study of normal development could provide a basis for a specification and hierarchical ordering of important features. Both of these approaches were recognised, however, as requiring contributions towards the better definition of the dimensions from both basic auditory processing and phonological constraints.

References

Carlson, R. and Granström, B. (1982) Towards an auditory spectrograph, in R. Carlson and B. Granström (Eds) *The Representation of Speech in the Peripheral Auditory System* Amsterdam; Elsevier Biomedical Press, 109-114.

Fourcin, A.J. (1977) Acoustic patterns and speech acquisition, in N. Waterson and C. Snow (Eds) *The Development of Communication.* London: J. Wiley, 144-171.

Fourcin, A.J. (1983) Perception of speech pattern elements, in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences,* Dordrecht: Foris 217-226.

Fourcin, A.J., Douek, E.E., Moore, B.C.J., Rosen S., Walliker, J.R. Howard, D.M., Abberton, E. and Frampton, S. (1983 Speech perception with promontory stimulation. *Annals of the New York Academy of Sciences, Vol. 405.* New York: New York Academy of Sciences 280-294.

Goldstein, J.L. (1983) An outline of recent research progress on fundamental pitch perception, in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences* Dordrecht: Foris, 245-247.

Goldstein, J.L. and Srulovicz, P. (1983) A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *Journal of the Acoustical Society of America* , 73, 1266-1276.

Kiang, N. Y-S (1965) Discharge patterns of single fibers in the cat's auditory nerve. *MIT Research Monograph no. 35.* Cambridge, Mass: MIT Press.

Patterson, R.D. Nimmo-Smith, I., Weber, D.L. and Miroy, R. (1982) The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America,* 72, 1788-1803.

Pickett, J.M. and Revoile, S.G. (1983) Speech cue use by the hearing-impaired, in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences* Dordrecht: Foris, 233-237.

Pols, L.C.W. and Stoop, K. (1983) Confusions between Dutch consonants under various conditions of noise and reverberation. in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences.* Dordrecht: Foris, 515.

Risberg, A. (1983) Frequency discrimination and speech perception in the hearing impaired, in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences.* Dordrecht: Foris, 227-232.

Ritsma, R.J. and Horst, J.W. (1983) Speech perception and frequency selectivity, in A. Cohen and M.P.R. van den Broecke (Eds) *Abstracts of the Tenth International Congress of Phonetic Sciences.* Dordrecht: Foris, 239-244.

Sachs, M.B. , Young, E.D. and Millerm M.I. (1982) Encoding of speech features in the auditory nerve, in R. Carlson and B. Granström (Eds) *The Representation of Speech in the Peripheral Auditory System,* Amsterdam: Elsevier Biomedical Press, 115-130.