



SPEECH-PATTERN PRESENTATION TO THE DEAF: SPEECH PERCEPTION AND PRODUCTION

A Faulkner and AJ Fourcin (EPI Group)

Department of Phonetics and Linguistics, University College London.
Wolfson House, 4 Stephenson Way, LONDON NW1 2HE, UK.

ABSTRACT

The optimal use of the impaired auditory system in the essential task of lipreading depends upon the transmission of invisible speech factors such as voice fundamental frequency, voicing, and manner of articulation. The same factors are also important for voice control. The corresponding speech features are imperfectly perceived by the deaf, but their acoustic correlates can now be relatively efficiently extracted in real-time (eg, IS Howard and JR Walliker: this meeting) and presented as an ensemble of simplified speech pattern elements matched to the patient's auditory abilities. Our recent receptive results show that combined pattern elements of voice fundamental frequency and voiceless friction can give additional benefit over voice fundamental frequency alone, and can also provide more benefit than whole speech presentation in a hitherto neglected sector of the deaf population. Matched speech pattern element presentation can also greatly assist speech production. The paper also examines the use of overall amplitude information in addition to voice fundamental frequency and voiceless friction.

SPEECH PATTERN PROCESSING HEARING AIDS

The whole speech signal is rich in acoustic information, to a degree which considerably exceeds the information processing capacity of the profoundly impaired auditory system. The idea of reducing this information processing load by the presentation to the deaf of simplified speech pattern elements has now a fairly long history [1]. Early informal experiments showed advantage from the simple use of an electrolaryngograph, and subsequent quantitative results made use of the voice fundamental frequency (F₀) pattern as an acoustic aid to lipreading, [2,3]. The principle of simplification was adopted in major cochlear implant research programmes in UCL and also in Melbourne. Recent advances in our understanding of the auditory abilities of the profoundly deaf have also led to developments in acoustic speech pattern hearing aids. The future for this work is now transformed by the prospect of computationally powerful digital processors of sufficiently small size and electrical power consumption to be applicable to wearable hearing aids, and by developments in speech signal processing which offer the potential for robust speech pattern element extraction.

SELECTION OF SPEECH PATTERN ELEMENTS

In different patients and environmental circumstances, different sets of features may be appropriate. This suggests a family of speech pattern element extracting aids which can provide the deaf with a smaller or larger ensemble of pattern elements depending upon their receptive capacity and upon the communicative situation, the language used, and the acoustic environment. Similarly, different sets of features would be appropriate at different stages of the language development of the deaf child or prelinguistically

deafened adult. In the provision of minimal auditory support to the adult lipreader, the essential speech pattern elements should correspond to invisible aspects of speech, such as voicing and prosodic contrasts, the presence of voiceless excitation, and the amplitude of vocal excitation. These speech elements, related as they are to features of voicing and manner of articulation, complement the visual cues which principally indicate place of articulation. The provision of voice productive feedback to the deaf speaker incorporating these features of vocal excitation and voice quality is also of considerable value by the enhancement of productive communicative ability.

AUDITORY ABILITIES OF THE DEAF

The auditory receptive abilities found in electrical stimulation of the totally deaf and acoustic stimulation of the profoundly deaf show many common features between the two groups of subjects, although there are also important differences [4]. Both groups respond primarily to low frequencies; in particular, thresholds are lower and the usable dynamic range is typically much larger at frequencies close to 100 Hz than at 500 Hz and above.

AUDITORY FREQUENCY ANALYSIS

In both electro-cochlear stimulation and the acoustic stimulation of residual hearing, the frequency analytic abilities of the normal auditory system are greatly impaired or absent. Although multiple electrode cochlear implants can provide some patients with a very limited stimulation of frequency selectivity, and some profoundly hearing impaired patients exhibit small but measurable selectivity [5], spectral processing abilities are minimal.

AUDITORY TEMPORAL PROCESSING

Unlike spectral processing, temporal processing is often relatively unimpaired in these listeners. For example, with temporally simple stimuli such as sinusoids, these deaf populations can show close to normal pitch discrimination for frequencies of 125 Hz. When they are presented with complex stimuli such as speech however, their ability to detect differences in voice fundamental frequency is often too limited to permit the reception of prosodic contrasts [5,6]. The detection of temporal gaps is somewhat impaired, especially if measured with noise [4], but when sinusoidal stimuli are used rather than noise, gaps of 5 to 20 msec can readily be detected both by implant users [7], and by our acoustically stimulated listeners. Sounds are also discriminable as periodic or aperiodic, typically down to duration thresholds of 50 ms or less. Finally, our recent measurements of the detection of sinusoidal amplitude modulation have shown detection thresholds of 10% or less for amplitude modulation of sinusoids at modulation rates of up to 40 Hz, which is close to the fastest modulation rate of speech amplitude envelopes.

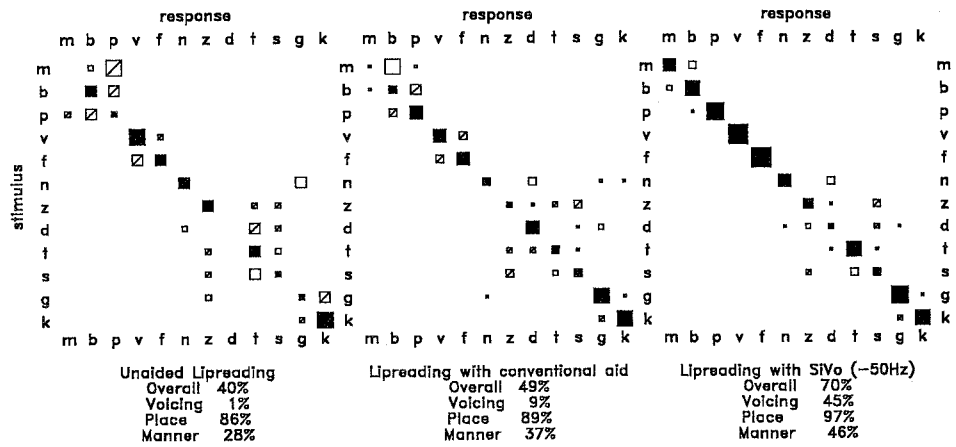


Figure 1. Intervocalic consonant confusions in aided lipreading. Left panel, unaided lipreading; centre, aided by the whole speech signal; right, aided by the voice fundamental frequency pattern from a SiVo aid, using the MAPITCH facility. The size of the sides of the squares indicate the proportion of a given response. Correct responses are shown as filled squares and voicing errors by a diagonal line. The figure also shows the overall percentage correct score and information transfer measures for features of place, voicing, and manner.

THE SPEECH CODING POSSIBILITIES.

These temporal processing abilities naturally lend themselves to the reception of essential invisible speech features relating to voicing and voice fundamental frequency, to the voiced/voiceless contrast, and to the amplitude of excitation. Sensitivity to temporal modulation also presents the possibility of the temporal coding of timbre. In as far as spectral processing remains or can be simulated, this can provide a natural means for the place coding of speech elements related to timbre.

OPTIMAL MATCHING AND CODING

An essential principle in the implementation of speech pattern element lipreading aids arises from the process of normalisation. This perceptual process has prepared the auditory system for the reception of stimuli having phonologically similar patterns but very different physical forms. Thus, speech pattern elements may be optimally matched to the auditory area through acoustic transformations which preserve linguistic structure. In this way, essential speech information can always be audible without discomfort; this is not readily achievable with conventional spectral shaping and compression techniques. For example, the EPI group's voice fundamental frequency based hearing aids have used the MAPITCH principle [1,3] by which the voice fundamental frequency pattern of a woman or child can be presented in a lower frequency region where the residual auditory abilities are less impaired, and with an expanded logarithmic frequency range. Similar considerations arise for the presentation of aperiodic patterns representing voiceless energy. These can be presented as low frequency sounds which retain the linguistically essential aperiodic aspect.

BENEFITS TO THE DEAF

The SiVo aid [3, 6] is a wearable voice fundamental frequency extracting aid which presents the voice fundamental frequency pattern as a sinusoidal acoustic stimulus of essentially constant loudness. It provides voicing and prosodic information with ideal control of intensity at any frequency, matched to the auditory area of the profoundly deaf lipreader.

SPEECH PERCEPTUAL RESULTS

Results from a group of eight post-lingually profoundly deafened adults [5] comparing the SiVo aid to excellent conventional aids have shown encouraging results. The subjects showed generally superior perception of prosodic contrasts when using the SiVo aid compared to the reception of the whole speech signal through a conventional hearing aid. In the aided lipreading of intervocalic consonants, two of the subjects were able to extract significantly more voicing information from the

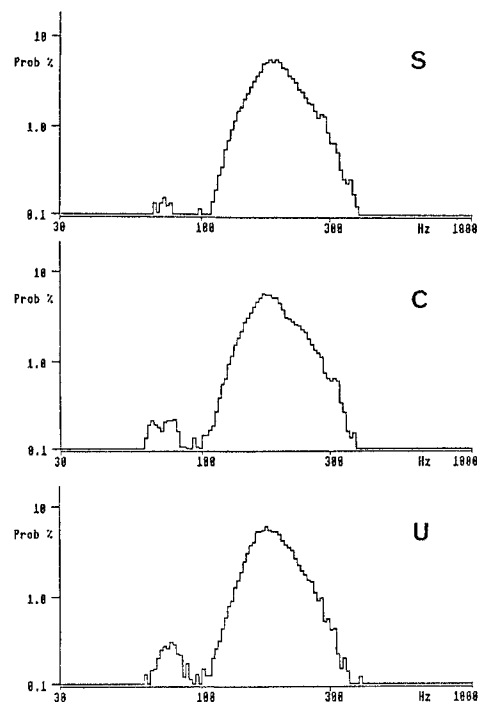


Figure 2. Voice fundamental frequency (F_x) distributions from a profoundly deaf female user of the SiVo aid. The distribution in the top panel (S) was produced while wearing the SiVo aid, the centre panel (C) while wearing a conventional hearing aid, and the bottom panel unaided (U).

SiVo aid than from the whole speech signal. Only one subject showed poorer reception of voicing information using the SiVo aid. Only two of the eight subjects were able to obtain useful information about fricative and other voiceless aspects of speech from the whole speech signal. Figure 1 shows aided lipreading results from one patient.

SPEECH PRODUCTIVE RESULTS

The speech production of the subjects who took part in the SiVo aid trial is presently undergoing analysis. Particular attention is being given to voice quality and fundamental frequency range, since these aspects are typically the most severely affected by post-lingually acquired profound hearing loss, and are readily quantifiable through study of the fundamental frequency distribution (Dx). Comparisons have been made of the effects of voice productive feedback from the SiVo aid and from a conventional hearing aid as compared to the distribution without feedback. Present results [8] show beneficial effects of feedback particularly in the tendency of the mode and range of the distribution to be more appropriate to the speaker's sex, and in the reduction of low frequency irregularity. An example of these Dx data from one patient are shown in figure 2. For this patient, feedback affects both the modal Fx and irregularity. Fx is most appropriate for this female speaker when she is using the SiVo aid. Most markedly, a very clear reduction in low frequency irregularity is seen when using the SiVo aid. A considerably smaller benefit is seen when the subject uses her conventional hearing aid.

ADDITIONAL SPEECH PATTERN ELEMENTS

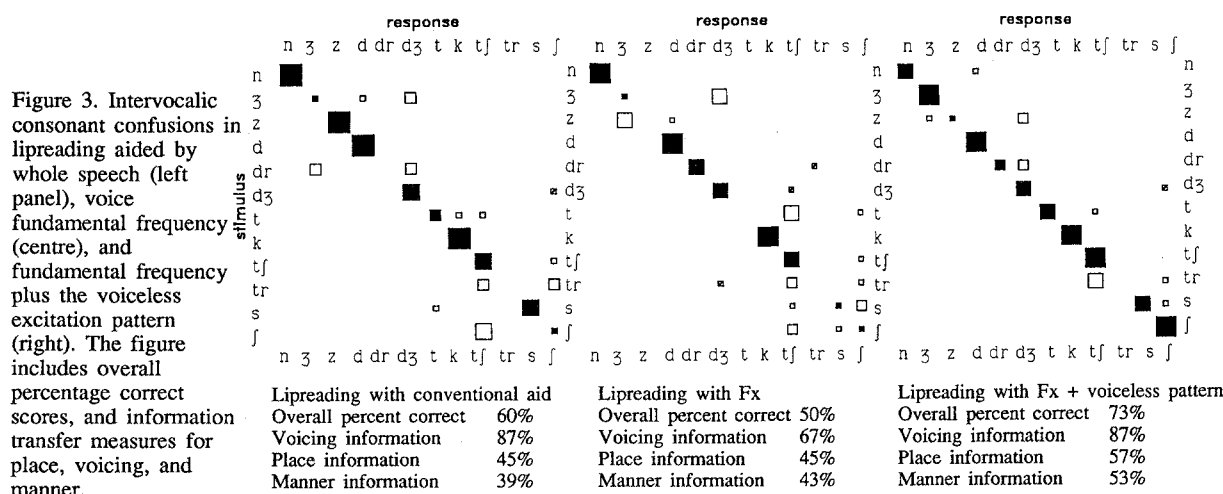
We have shown that the addition of an aperiodic auditory pattern to represent voiceless speech has significant additional receptive benefits over voice fundamental frequency alone [5,9]; figure 3 illustrates results obtained from one profoundly deaf subject. Next we shall consider the additional advantages of a matched representation of the overall speech amplitude envelope, and also consider the use of speech pattern presentation in languages other than English. In addition to the receptive advantage, amplitude information is also likely to be of considerable significance in the control of the deaf speaker's own voice.

The greatest information load in speech is carried by consonants, and since in the main European languages, consonants are most often intervocalic, we will concentrate upon speech pattern representations of VCVs. In lipreading English, VCVs can be partially distinguished simply on the basis of the voice fundamental frequency pattern. The alveolar consonants [d n t s] are illustrated in figure 4 together with voice fundamental frequency patterns, and with compound patterns which also represent voiceless excitation and overall amplitude. The voice fundamental frequency pattern can clearly contrast voiced against voiceless consonants, particularly since the unvoiced portion of the voiceless sounds is rather long relative to the gap thresholds found in the deaf. The contrast between voiced plosives and nasals can also be represented when, as frequently occurs, a voiced plosive shows brief devoicing. More robustly, this contrast can be signalled from the compound pattern by the different amplitude envelopes of the nasal and the voiced plosive. The addition of a cue to voiceless excitation can additionally contrast [s] and [t], and in other homorganic consonant sets, this pattern together with voice fundamental frequency can convey the fricative/affricate contrast.

The significance of these speech pattern elements will be different in different languages. For example, in French, devoicing will be unlikely in voiced plosives, so that the fundamental frequency pattern will not contrast nasals and voiced plosives; hence, amplitude envelope will be a more important factor than in English. French also differs from English in that the French voiceless plosives are not strongly aspirated, so that the addition of a voiceless pattern may be of particular salience in the French plosive/fricative contrast.

NEW PATTERN EXTRACTION METHODS

Advances in speech processing are now promising considerable developments in robust pattern element extraction. For example, neural-net methods have provided excellent results in time-domain fundamental period extraction [10], and these algorithms also have potential as a basis for robust voiceless detection and formant frequency estimation. Our colleagues have demonstrated [11] that these methods can be implemented on standard CMOS digital signal processing devices to operate in real time with reasonable power consumption, and we are now developing wearable speech pattern extracting hearing aids using these algorithms and processors.



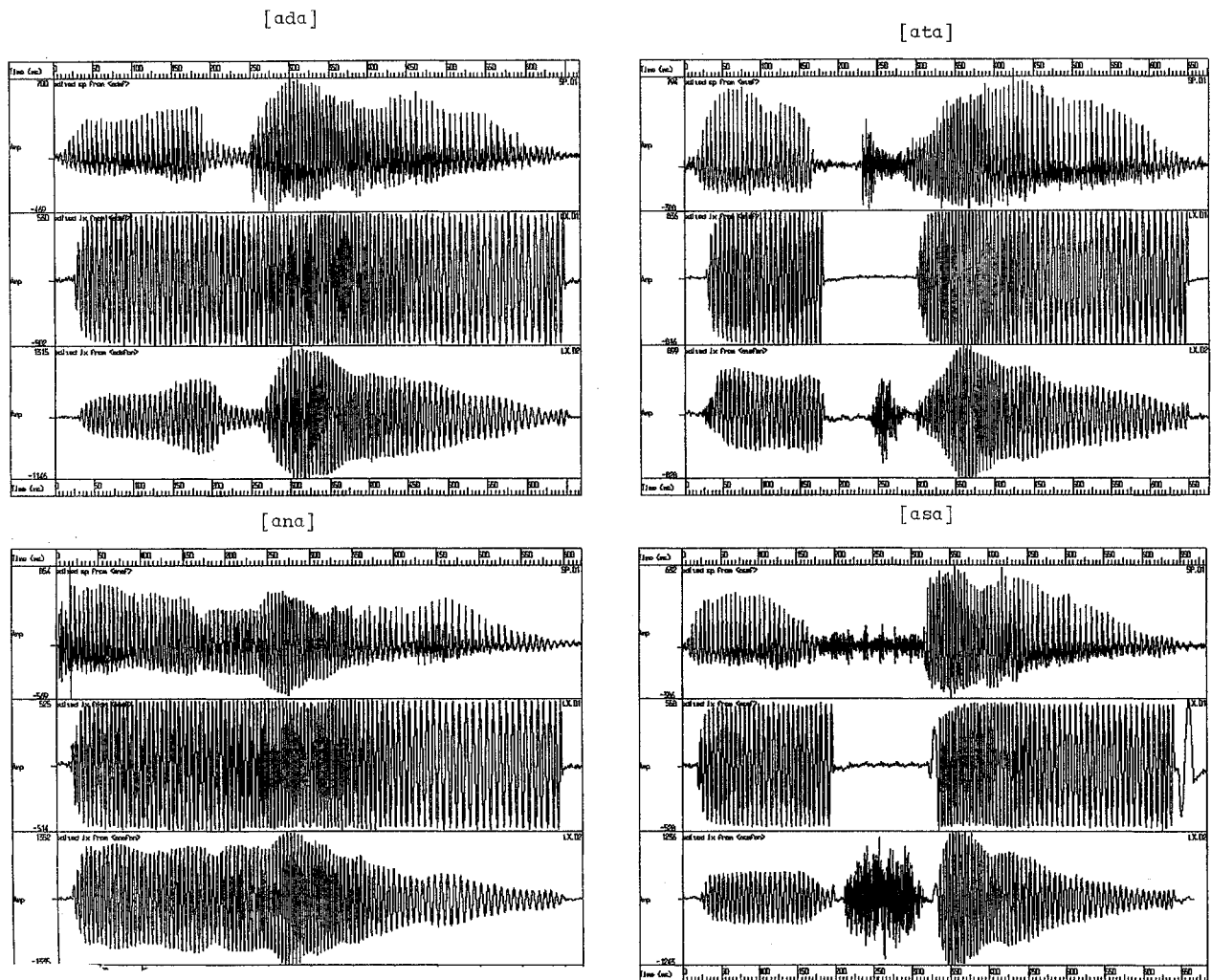


Figure 4. Intervocalic alveolar British English consonants in an [a] vowel frame. Upper left, [d]; lower left, [n]; upper right, [t]; lower right [s]. In each panel the top plot in the speech waveform, the central plot is the sinusoidal voice fundamental frequency pattern as from the SiVo aid, and the bottom plot is the compound fundamental frequency, voiceless excitation, and overall amplitude pattern.

ACKNOWLEDGEMENTS

This work is supported by the Medical Research Council (UK). Our colleagues in the External Pattern Input group have all played important parts in the work reported here.

REFERENCES

- [1] AJ Fourcin, SM Rosen, BCJ Moore, EE Douek, GP Clarke, H Dodson, LH Bannister, "External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings". *Br. J. Audio.*, Vol 13, pp 85-107, 1979.
- [2] SM Rosen, AJ Fourcin, BCJ Moore, "Voice pitch as an aid to lipreading" *Nature*, Vol 291, pp 150-152, 1981.
- [3] A Fourcin, E Douek, B Moore, E Abberton, S Rosen, J Walliker, "Speech pattern element stimulation in electrical hearing", *Arch. Otolaryngol.*, Vol 110, pp 145-153, 1983.
- [4] S Rosen, A Faulkner, DAJ Smith, "The psychoacoustics of profound hearing impairment", *Acta Otolaryngol.*, in press.
- [5] A Faulkner, AJ Fourcin, BCJ Moore, "Psychoacoustic aspects of speech pattern coding for the deaf", *Acta Otolaryngol.*, in press.
- [6] S Rosen, JR Walliker, AJ Fourcin, V Ball, "A microprocessor-based acoustic hearing aid for the profoundly impaired listener", *J. Rehab. Res. and Devt.*, Vol 24, pp 239-260, 1987.
- [7] BCJ Moore, BR Glasberg, "Gap detection with sinusoids and noise in normal, impaired and electrically stimulated ears", *J. Acoust. Soc. Am.*, Vol 83, pp 1093-1101, 1988.
- [8] G Ball, A Faulkner, "Speech production of postlingually deafened adults using electrical and acoustic speech pattern prostheses" *Speech, Hearing and Language: Work in progress*. University College London. Vol 3, pp 13-32, 1989.
- [9] A Faulkner, C Potter, G Ball, S Rosen, "Audio-visual speech perception of intervocalic consonants with auditory voicing and voiced/voiceless speech pattern presentations" *Speech, Hearing and Language: Work in progress*, University College London. Vol. 3, pp 87-106, 1989.
- [10] IS Howard, MA Huckvale, "Speech fundamental period estimation using a trainable pattern classifier", *Proc. Speech '88: 7th FASE Symposium*. Inst. Acoust., Edinburgh, pp 129-136, 1988.
- [11] JR Walliker, IS Howard "The implementation of a real time speech fundamental period algorithm using multi-layer perceptrons" *Speech Communication*, in press: (paper given at Neurospeech, Edinburgh, May 1989).