1. Introduction. In this paper we consider a framework against which to discuss strategies for the design of speech processors for cochlear implant prostheses. We hope to encourage discussion of the bases for such a framework even though it may seem a distant objective owing to the large gaps in our understanding of several component parts of cochlear implant systems. The existence of such a framework would provide a background against which to view the current diverse cochlear implant systems and to evaluate their performance.

2. A Cochlear Implant System. We divide a cochlear implant system into four component parts: external speech processing; signal transfer; neural interface; and the perceptual system of the patient. These four components are clearly quite diverse in character yet they are tightly bound in the overall design and operation of the total prosthetic system. The most fundamental function of the system is the ability to generate neural discharge patterns on the auditory nerve which may be controlled by external signals. These signals must be derived from the acoustic speech signal by a speech processor, and then be transferred to electrodes at the neural interface. Once this signal pathway has been established, the perceptual domain may be explored.

2.1 The Neural Interface. The physical and neurophysiological condition of the cochlea, together with the configuration and siting of electrodes comprise the most salient aspects of the neural interface. The placement of electrodes and the capabilities of the driving electronics may be selected according to a combination of knowledge about the physiology of the mammalian cochlea, the acoustic cues used by normal hearers in perceiving speech, and previous studies of electrical stimulation of animals or patients. Considerable differences have been noted in neural population among profoundly deaf persons (Hinojosa & Marion, 1982), and in the approaches of various research groups to electrode implementation and siting (Eddington, 1980). The performance of the system at the neural interface must therefore be regarded as highly idiosyncratic and therefore uncertain unless it can be established by means of feedback from the electrically induced activity in the peripheral auditory system.

2.2 Signal Transfer. Options for the implementation of the signal transfer component of the system include direct linkage of current drivers to the implanted electrodes via a percutaneous plug, or a transcutaneous radio linkage to the electrodes, or to an implanted receiver and current stimulator unit. The characteristics of the transfer system define the precision with which the intended current waveform may be implemented. An analogue channel is susceptible to the uncertainty of additive noise, but provides maximum precision. Feedback of inter-electrode impedances may also be achieved if a percutaneous approach is used. A digital channel largely avoids transmission noise problems but introduces limits in precision in both time and signal strength.

2.3 Speech Processing. The speech processor, being totally external, is the focus of fine tuning that is necessary to transform a theoretically conceived system into an operational one. The forms that this component of the system can take are numerous but are constrained by such factors as the quality of the incoming speech signal, the implementation of the neural interface, and the mode of signal transfer. These factors will dictate the maximum number of channels on which independent information may flow, the temporal and quantitative precision of each channel, and the susceptibility of the system to noise. Within these constraints many options remain including the number and content of independent channels, and the specific transformation of that content into a signal that can control the flow of current at the neural interface.

2.4 The Perceptual Mechanism. The perceptual mechanism that is employed by the patient in responding to the electrical stimulation of his auditory nerve must be considered in any attempt to set criteria for the design of speech processing. There is evidence to suggest that human ability to process new types of input through the auditory system is relatively inflexible beyond the...
age of puberty (Lenneberg, 1967; Tahta et al., 1981). If this is generally true it would be very important to extensively assess the status of the auditory system of the patient before deciding on a speech processing strategy. This would imply the use of a large battery of psychophysical tests via the electrical stimulation channel, to explore the assumed invariant perceptual system. If however the adult post-lingually deaf person can learn to analyse a new set of auditory inputs and integrate them into his speech perception process, then emphasis would be placed on training at a speech perception task level. The experiment of Blessen (1972) provides evidence that flexibility to use a novel auditory channel varies over individual subjects. He placed a spectral inverter circuit in an audio link between pairs of adults. Their task was to establish and develop speech communication via the link alone. After a period of 15 hours using the link some subject-pairs succeeded in adjusting to the new auditory input, but others did not. Performance in conversation did not correlate with ability to decode transformed spectral cues as measured by word or phoneme identification scores. Subjects-pairs who communicated well appeared to be those who were able to integrate prosodic features of pitch, stress, and temporal patterning, which were unaffected by the spectral inversion, with syntactic and semantic structural knowledge (Stark, 1974).

Knowledge of the potential of a cochlear implant patient for processing highly unfamiliar auditory input would be most useful information, both in the process of selecting patients and for designing speech rehabilitation. The differences between subjects in Blessen's experiments may be related to inability in some subjects to utilise new cues for percepts, or to inability to rely heavily on structural linguistic knowledge, or perhaps to motivational differences. Amongst the cochlear implant patients we may presume a very high degree of motivation to perceive speech, but in addition to any differences in perceptual flexibility, each patient may be receiving different information depending on the status of their neural interface.

An informal survey of case histories of cochlear implant patients would indicate that some learning, indicated by improved scores on tests separated by a period of months, does take place for adults (e.g. Thielemeier et al.). The basis of such improved performances has not been reported.

2.5 Summary. It is clearly difficult to compare the performance of one cochlear implant system with another if the total differences between the two systems cannot be adequately defined or controlled. Similarity and difference may be quite easily defined for the speech processor and signal transfer components. The neural interface may only be so defined if extensive electrical testing is performed using feedback of the induced neural activity. The perceptual systems that receive the neural excitation may be described in very broad terms such as the age of the patient, the period of deafness, and the duration and type of training with electrical stimulation. This means that valid comparisons between the performance of different systems may only be made on a statistical basis over groups of implanted patients selected according to gross similarity of all components of their implant system save the one on which comparison is based.

In the rest of this paper we wish to consider the basis of a model for the selection of a speech processor given that a "standard" form of processor is unlikely to be optimum for many patients.

3. ASPECTS OF A GENERAL MODEL. A useful distinction that has been drawn in considering the processing of speech and language both in man and machine is that of "bottom-up" versus "top-down" processing (e.g. Ainsworth, 1976). A "bottom-up" process examines the raw evidence and extracts information that is required by some higher stage of processing. A hierarchy of such processes may be constructed to comprise a total system that generates the required output at some high level. A "top-down" process generates a hypothesis which it tests against a range of inputs which may be the outputs of other processes. It can be resistant to distortions of these inputs as it can be selective and adaptive in the attention that it pays to each input.

To some extent all speech processing systems have both "top-down" and "bottom-up" aspects even if one or the other is very minimal in extent and is restricted to the extremities of such a system. In the most extreme top-down model a potential message is hypothesised on the basis of linguistic constraints
and the acoustic signal that would correspond to this message is tested against the incoming acoustic signal. A succession of such tests may be made for a range of potential messages until a match is achieved. In the most extreme bottom-up model the acoustic signal is analysed progressively such that a potential message or set of potential messages is obtained. The comparison is now made among these messages in terms of their lexical, syntactic and semantic plausibility.

In normal speech perception the interface between top-down and bottom-up processes may be considered to vary dynamically. In a highly restricted linguistic situation where semantic, syntactic and lexical constraints are well defined, a top-down approach to a fairly low level may be most efficient. A similar approach may be efficient if the speech signal is of poor quality, as bottom-up analysis of a poor quality signal can lead to uncertain higher order analyses. Conversely if the signal quality is very good and the linguistic constraints are relatively weak, then a bottom-up approach to a high level may be the most efficient approach.

We now relate this theoretical idea of dynamically interacting bottom-up and top-down processes to the design of a cochlear implant system. We can postulate that the purpose of the fine tuning at the level of the speech processor, assuming a well designed and implemented neural interface, is to enable communication between the upward-flowing input signal information and the downward-reaching cognitive processes to be consistently achieved. The selection of a speech processing option will influence the information pathways between the electrical input and the speech perceptual output of the total system.

Lack of understanding about the detailed form, or "wiring diagram", of the ear-brain-speech-perception system requires that we consider it to be a "black box". The classical "black box" has inputs and outputs, and internal states that cannot be measured. The functional relationships of outputs and inputs must therefore be determined empirically. In a cochlear implant system the inputs are determined by the nature of the speech processor, signal transfer hardware, and neural interface of the device. Potential outputs exist in several different domains and cannot be related directly to positions along the information pathway, but do represent physiological or perceptual activity that is correlated with the input. Direct physiological measurements using invasive techniques such as single unit neuronal potential recording at various points along the auditory pathway are possible only in parallel studies in mammalian auditory systems. Non-invasive physiological measures such as evoked potentials which can sometimes be related to auditory processing may be performed on human patients (Chouard, 1978; Greenberg, 1980). The major outputs of the system are however the wide range of different perceptual responses from the patient. Although we may establish correlations between outputs and inputs and between the outputs themselves we do not know in any detail the precise ordering of processes that underlie what we are able to measure (Wood, 1974).

For the post-lingually deaf patient any auditory or linguistic percept that is shared by the normal hearing population and the patient may be used as an output. The designer of the speech processor can use these outputs as feedback to optimise the external speech signal processing such that the detection, acceptability, or discrimination of the particular input is improved.

In summary, a theoretical framework against which different speech processing strategies may be viewed is that of the set of "black-box outputs" that is used to determine the best form of speech processing. Three distinct categories of black-box output may be identified: the physiological, the psychophysical, and the linguistic. Clearly all system designers utilise certain physiological knowledge to ensure long-term biocompatibility, psychophysical feedback to ensure acceptance of signals by the patient, and linguistic feedback to examine the performance of the system once the speech processor has been selected. Outputs used for these purposes do not form part of the proposed theoretical classification. It is based solely on outputs that are used as feedback variables in the optimisation of the processor.

3.1 PHYSIOLOGICAL ASPECTS. One theoretical model often advanced in the literature on electrical stimulation of the auditory system is a physiological simulation model (Kiang & Moxon, 1972; Evans, 1977; Clark et al., 1978; Kiang
et al., 1979; Merzenich & White, 1980). This model entails the matching of physiological observations during electrical stimulation of an implanted ear with those observed during acoustic stimulation of a normal ear. Such observations may be the discharge rate of auditory nerve fibres (Sachs & Young, 1979), the temporal pattern of such discharges (Young & Sachs, 1979), or maybe the discharge of pattern-specific higher order neurones in the cochlear nucleus (Palmer & Evans, 1979) or auditory cortex (Whitfield & Evans, 1965).

There are major barriers to achieving a satisfactory simulation of normal activity in these parts of the auditory system. Firstly, a simple transformation between acoustic and electrical stimulation does not produce equivalent neural discharge patterns (Kiang & Moxon, 1972). Substantial advances are required in the modelling of the electrical stimulation process, and in the means of controlling current flow in the cochlea, to produce neural activity patterns which match the observed activity of the acoustically stimulated cochlear nerve. Secondly, it has been shown that certain cells in the cochlear nucleus are sensitive to acoustic patterns whose representation is not evident in simple measurements of the more peripheral neural activity in the cochlear nerve (Palmer & Evans, 1979). Not only is there a lack of ability to synthesise appropriate inputs to a phonetic stage of processing, but the nature of the patterns necessary to permit upward propagation of auditory information in the normal manner is not fully understood. It seems highly likely that a prosthetic system which does not provide the fine detail of neural information necessary to trigger such feature detectors will suffer restrictions on the upward propagation of information. Although speech is intelligible under many gross distortions there is evidence that small auditory details can be very important in facilitating perception performance (Millar & Ainsworth, 1972).

3.2 PSYCHOPHYSICAL ASPECTS. The essence of a psychophysical model is to link acoustic and electrical stimuli on the basis of their similar psychophysical characteristics. A speech processor based on this type of model transforms the acoustic signal of speech into electrical stimuli which elicit responses which have psychophysical similarity to those which are elicited by acoustic stimulation of the normal ear. The use of psychophysical outputs from the perceptual black-box as a basis for developing speech processing assumes that a direct or indirect mapping from psychophysics to speech perception exists.

Wood (1974) cites evidence that the psychoacoustic percept of pitch and the phonetic percept of place of articulation are at least partially mediated by parallel perceptual processes. This evidence excludes the possibility of a totally serial model in which all the results of auditory processing may be assumed to be potential inputs to a phonetic stage of processing. However, a model which assumes that the processing underlying the percepts of loudness and pitch is common to phonetic processing has been applied to the modelling of speech acoustics for automatic speech recognition (Zwicker et al., 1979), in reducing the perceptual impact of noise during the presence of speech (Schroeder et al., 1979), in the clinical rehabilitation of the hearing-impaired (Asp, 1973), and in designing a simple vocoder (Zollner, 1979). There is however no general agreement that sound attributes such as loudness and pitch are the primary perceptual dimensions of speech. It appears quite feasible, however, that certain speech entities, such as words, could have auditory profiles which may be used as alternative foundations for speech processing. Furthermore, complex attributes such as voice-likeness and naturalness which are more difficult to quantify may be significant.

3.3 SPEECH PERCEPTION ASPECTS. In a speech perception model the nature of the electrical signal to transmit the speech information is determined with minimal reference to the physiological or psychophysical features. Attention is focussed on presenting as much information as possible about the incoming acoustic signal, and relying on the integration of the new auditory channel into a perceptual system comprising all speech information channels. A successful speech processing strategy will be one that demonstrates integration of such multiple channel information. This is attractive because it places weight on the top-down processes of speech perception which would be least affected by the new auditory channel. Support for this model may be generated by consideration of the natural process of speech and language acquisition. A child will develop
speech and language skills when placed in an environment where an adequate number of acoustic cues are available and where speech and language have a contextual relevance in life's goals. It is, however, a matter of dispute whether the process involved is based on adaptation to the auditory environment (Simon & Fourcin, 1978) or is based on invariant analysis structures within the auditory system (Eimas, 1971; Blumstein & Stevens, 1979). The appeal to the use of intact top-down processing abilities may be of little consequence considering the possible lack of auditory flexibility in some adults. If psychological evidence of the flexibility of the individual patient could be obtained such an approach may however be particularly appropriate.

4. SUMMARY & CONCLUSIONS. This paper has proposed a rudimentary theoretical framework for the description of speech processing strategies for cochlear implant prostheses. Description based on physiological, psychophysical, and speech perceptual feedback have been suggested as the major definitive dimensions of this framework.

The optimum speech processor will be that which generates black-box outputs which match natural hearing physiologically, psychologically, and in speech perception performance. Just as a less than optimum speech synthesis system can produce acoustic signals which are useful for communication between a machine and a person, so a less than optimum speech processor will be able to produce electrical signals which are useful for communication between a speaker and a cochlear implant patient. It is necessary to discover, however, which of our postulated dimensions of speech processor design have the greatest potential for further development. What advances in knowledge, technology, and experimental ingenuity are required to realise such development?

Baltimore: University Park Press.


Author/s:
Millar, J. B.; Tong, Y. C.; Clark, Graeme M.

Title:
Selection of speech processing for cochlear implant prostheses

Date:
1983

Citation:

Persistent Link:
http://hdl.handle.net/11343/27184

File Description:
Selection of speech processing for cochlear implant prostheses

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.