AUDITORY NERVE ARRAY REPRESENTATION OF COMPLEX ELECTRICAL AND SOUND STIMULI.

Michael M. Merzenich, Departments of Otolaryngology and Physiology, University of California, San Francisco, California 94143.

I. INTRODUCTION

There have been a number of described strategies for encoding sound in a multichannel electrical stimulation auditory prosthesis. These strategies have followed two basic lines. The UCSF group has adopted what has been termed a "limited physiological simulation" strategy. The objective is to directly simulate essential features of the normal spatio-temporal neural representations of sound, through appropriate spatially and temporally patterned stimulation of the auditory nerve array (1-3). The simulation of physiological defined response patterns is "limited", in the sense that: a) studies indicate that it is probably essential to represent only specific aspects of the total distributed neural representational pattern; b) representation of non-essential aspects of response patterns gives rise to distortion; and c) simulation of details of normal representational patterns of complex stimuli (e.g., speech) is necessarily crude, in some (but not all) ways.

Another class of physiological simulator has been described recently by Kiang and colleagues (4). It has been implemented in a highly reduced form (again effecting a "limited simulation") in the patient studies of Eddington (5).

The second basic approach could be termed a "perceptual simulation" coding strategy. It has involved an attempt to translate simple perceptual responses in the pitch and loudness domains, by addition, into recognition or discrimination in the domain of speech. Devices of this type have been developed by several implant groups, including those at Stanford (6), Melbourne (7), and Paris (8). Other groups have yet to describe their coding strategies, but at least most are clearly following a perceptual simulation model. Described results with the Austrian multichannel device are not comparable to other multichannel devices, as they employed a channel vocoder that also incorporates a major (several octave) frequency translation in coupling (9).

In this brief report, some of the bases of our rationale for development of a limited physiological simulation strategy shall be outlined. Some patient results constituting an initial evaluation of this strategy will then be described.

II. LIMITED PHYSIOLOGICAL SIMULATION STRATEGY

The starting point for effecting a physiological simulation is at the interface of the electrode array and the auditory nerve array, for any effectively patterned simulation requires appropriately controlled nerve array excitation and a predictability of spatial and temporal features of evoked neural responses. By contrast, in perceptual simulation, there is no such requirement, no necessity for any stimulating element to be in any particular location, or to necessarily control excitation in any particular way.

Through a long series of directed studies, we have developed a multielectrode array with which appropriately controlled sector-by-sector control of nerve array excitation can be effected. In order to gain such control in nerve sectors with good and poor nerve survival, it is necessary to employ bipolar stimulation, with stimulation electrodes positioned near the upper surfaces of the scala tympani (10,11). Control of excitation is degraded with use of monopolar electrodes (especially when nerve survival is poor), and with the use of common grounding schemes. It is also degraded if bipolar stimulating elements are moved away from the surfaces of the scala. Despite long awareness of these considerations, the UCSF group is still the only group that employs a bipolar array with stimulating elements positioned near the defined "ideal" locations. Why might that be the case? First of all, many investigators including most not listed above are fundamentally employing a "what does that sound like" perceptual simulation strategy. The electrode elements are, again, of less consequence for such device development. Second, true multichannel bipolar driving devices are more difficult to implement electronically. Perhaps a third reason has been a general pessimism as to the faithfulness with which a physiological simulation can be effected. Obviously, if essential aspects of eighth nerve response patterns cannot be reproduced, then either a perceptual simulation, or a representation of speech in the most information-rich way with no attempt at effecting recognition would probably constitute the favored strategies.

What, then, in a physiological model must be simulated? And what can be simulated with patterned electrical stimulation?

A. Representation of Speech By Normal Distributed Auditory Nerve Responses.

Recent studies by Sachs, Young and colleagues, and Kiang, Delgutte and colleagues have revealed the likely bases of the coded representation of speech elements across the auditory nerve array (12-15, 4). The studies of Young and Sachs have revealed that temporal features of response almost certainly represent different formants of consonants and vowels, and that the critical zones of representation are in the resonant regions of representation of these formants (12,13). This is consistent with earlier evidence that spectral information is principally coded temporally (e.g., see 16, 17). What Sachs, Young and colleagues have shown is that the auditory nerve fiber discharge rate vs cochlear place profiles do not likely constitute the basis of coding of speech elements, as above 50-60 dB SPL rates are saturated over a broad cochlear sector (18,19), and do not separately represent different vowel formants. If phase-locking of response at the best frequency at each site is defined as a function of place, a highly robust (as a function of stimulus level, and in the presence of noise) representation of the components of the vowel or consonant sound emerge. This finding led Sachs and colleagues to propose that temporal information must be decoded place by place, an idea independently developed by the UCSF cochlear implant group from consideration of results of electrical stimulation in implanted patients (20).

It should be noted that, by this view, the speech element is represented by only a fraction of the total available information. What must be simulated is the representation of speech element components over sectors about 1½ to 5mm across. The physiological simulation is reduced to a representation of appropriate temporal information in the appropriate cochlear location, with control on the grain of a few millimeters absolutely required to implement the model. It should be stressed that it is not necessarily the stimulus period itself that constitutes the actual temporal code. On the basis of electrical stimulation results, we have argued that it is more likely the details of interfiber timing that are being decoded centrally (20,21).

B. How Faithfully Can These Representational Patterns Be Simulated?

In earlier studies, it has been demonstrated that two essential features of a physiological simulation of auditory nerve array response patterns can be effected. First of all, appropriate spatial control of stimulation of sectors 1½ to 2mm long has been achieved with "ideal" bipolar stimulating elements (10,11). Second, phase-locking generated by electrical stimulation can be equivalent to or superior to the normal (22,23). Thus, it is possible to deliver a temporal representation of the frequency components of speech elements to appropriate locations, with acceptable control of channel interactions. This has now been directly demonstrated to be the case in human subjects (24). In this vein, it should be pointed out that because the representations of components "overlap" in the normal case does not mean that they can be allowed to overlap in a cochlear prosthetic device. A physiological simulation coding device represents resonance location information presumably essential for the reproduction of features of speech response patterns that the central nervous system decodes, and overlap of that information (which does not overlap substantially normally) should result in perceptual distortion.

There are, of course, details of normal representation that cannot be faithfully reproduced with electrical stimulation. The distributed pattern of temporal information as a function of cochlear location is impossible to reproduce, as it is necessarily symmetrical. Idealization of this feature of response might be important because, again, it appears likely that the central nervous system actually decodes interfiber temporal information around each resonance location (20,21). It should also be recognized that control on a grain finer than about 1-2mm is not possible with intrascalar stimulation, i.e., no improvement in the fidelity of spatial representation can be gained by adding electrode elements (1,25), because of the necessary distances of stimulating elements from auditory nerve fibers as well as practical difficulties in mechanical design.

What are the advantages of this strategy? Why not pursue a more "complete" limited-channel-number nerve array response simulation model, as has been elegantly described by Kiang, Eddington, and colleagues (4). In fact, our starting point in simulation is electrode performance, and experimentally defined electrode performance argues against such model implementation (10, 26,27). Consider a simple case in point. At high sound levels in a normal ear, a single low frequency tone will excite neurons at maximum saturated firing rates across most of the cochlear nerve array (18,19). Electrical stimulation at any given cochlear site produces an electrode-specific pitch. Simultaneous stimulation at a series of sites (required for a complete simulation of this single tone's distributed neural response pattern) evokes a noise, and not a simple sensation. That occurs, it is believed, because any given electrode creates a local, detailed spatial distribution of temporal information that is decoded as a site-specific pitch. It has been found empirically that these electrode-specific qualities are difficult (probably impossible) to avoid. Fortunately, the studies of Sachs and colleagues and of others have suggested the pitch is "read" from temporal information at and near the resonant location. For these and other similar reasons, we believe that implementation of a complete simulation model is not practical.

At the present time, perceptual simulation model options (like, for example, the Stanford "vocal tract resonance" model or the Melbourne formant extracting model) are not being pursued by the UCSF group. Proper addressing on the output side of such a model should effect an almost certainly degraded version of the consequences of a channel vocoder like that employed in a limited physiological simulation device. The coupling of such a minimal coder with an inherent frequency translation (i.e., by coupling with regard to "perceptual simulation" but not cochlear place) is probably doomed to failure. It is no surprise, then, that only near-chance open speech recognition has been achieved with application of perceptual simulation multichannel models.

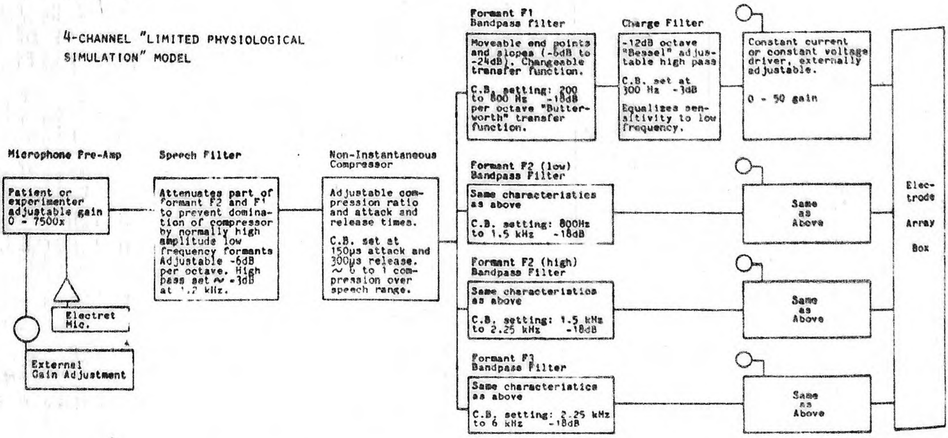
III. RELEVANT PATIENT RESULTS

The basic configuration of the multichannel speech processor model not employed experimentally is illustrated schematically in Figure 1. The compressor and filter settings for the last studied patient (C.B) are also indicated. Basic compressor and filter parameters were adjustable as indicated, in this analog realization of a 4-channel model. The wearable device for this patient was essentially identical, except that channel outputs are fed to an RF transmitter stage. In the device optimization strategy being used: 1) Loudness information derived in basic psychophysical studies provided the parametric basis for initially adjusting the model. Thus, as an example, threshold, growth of loudness and maximum comfortable loudness determinations were used to set compressor ratios and functions. 2) The consequences of manipulation of simple model parameters were evaluated, using special open two-choice single syllable item lists in which items differed in one vowel or consonant feature. Thus, for example, the compressor attack and release times were varied over a wide range to determine the optimal range of operation. (In this example, device performance was found to be greatly degraded when 5- to 10-fold increases in these values were made.)

With this 4-channel and with a 3-channel model, very significant levels of open speech recognition were recorded. To briefly summarize: 1) for 2-choice items, open list, vowel first formants different, the last 25 tests (200 items) with device in standard configuration (as in Figure 1): 99% correct. 2) 2-choice items, open list recognition, vowel first and higher formant different, last

FIGURE 1

4-CHANNEL "LIMITED PHYSIOLOGICAL SIMULATION" MODEL



25 tests, standard configuration: 100% correct. 3) 2-choice items, open list recognition, vowel second formants different, tests in which second formant range was devolved across 2 channels: 72% correct. Our benchmark 13-test Minimal Auditory Capacity series (26) was conducted entirely with the use of a 3- and not a 4-channel model (the 2nd formant range was represented by a single channel) in this patient. Some results include: 4) open 4-choice vowel recognition test, 128 items (25% chance): 74% correct. An analysis of errors revealed that most items missed involved misidentification of dipthongs as steady state vowels, or were errors dependent upon second and higher formant distinctions accountable by the simple fact that these components were at least usually not developed across output channels. 5) open 4-choice initial consonant features recognition, 128 items (25% chance): 89% correct. Above chance performance was recorded for all consonant feature distinctions.

These are the highest overall scores on these basic speech element recognition tests yet reported for a cochlear implant multichannel speech processor. The subject also recognized a significant percentage of words in sentence, of spondees, and of open single syllables. Improvements in speech recognition with use of multichannel as compared to single channel operation in this patient were dramatic and immediate (performance with single channel models was, for all tests, at near-chance levels). Further, device function was substantially degraded with use of monopolar as compared with bipolar stimulation. Other studies have revealed, again, that interelectrode interactions are significantly greater with monopolar than with bipolar stimulation (24). Finally, multichannel test results are directly interpretable in terms of a limited physiological simulation model, in that most open recognition test correct responses and errors are understandable in terms of the predicted devolution of speech elements across electrode array outputs.

ACKNOWLEDGEMENTS

Supported by NIH Contract NO1-NS-0-2337, NIH Grant NS-11804, Hearing Research, Inc., and the Coleman Fund. The author thanks A-C Guerin-Weaver and William Garrett for assistance in the preparation of this manuscript.

REFERENCES CITED

1. Merzenich, M.M. (1983) Coding of sound in a cochlear prosthesis. Some theoretical and practical considerations. Ann. N.Y. Acad. Sci. (In Press).
2. Merzenich, M.M., White, M., and Vivion, M.C. (1979) Some considerations of multichannel electrical stimulation of the auditory nerve in the profoundly deaf; interfacing electrode arrays with the auditory nerve array. Acta Otolaryngol. 87:196.
3. Merzenich, M.M., and White, M. (1980) Coding considerations in design of cochlear prostheses. Ann. Otol. 89, Supp. 74:84.

4. Kiang, N.Y.S., Eddington, D.K., and Delgutte, B. (1979) Fundamental considerations in designing auditory implants. Acta. Otolaryngol. 87:204-218.
5. Eddington, D.K. (1983) Results of intracochlear stimulation. Ann. N.Y. Acad. Sci. (In Press).
6. White, R.L. (1982) Review of current status of cochlear prostheses. IEEE Trans. Biomed. Eng. 29:233.
7. Tong, Y.C., Clark, G.M., Dowell, R.C., Martin, L.F.A., Seligman, P.M., and Patrick, J.F. (1981) A multiple channel cochlear implant and wearable speech-processor: An audiological evaluation. Acta Otolaryngol. 92:193.
8. Chouard, C.H., Fugain, C., Meyer, B., and Lacombe, H. (1983) Long term results of the multichannel cochlear implant. Ann. N.Y. Acad. Sci. (In Press).
9. Hochmair, E.S., and Hochmair-Desoyer, I.J. (1983) Percepts elicited by different speech coding strategies. Ann. N.Y. Acad. Sci. (In Press).
10. Merzenich, M.M., and White, M. (1977) Cochlear implant. The interface problem. In: Functional Electrical Stimulation. T. Hambrecht, and J. Reswick, eds. Marcel Dekker, N.Y.
11. Vivion, M.C., Merzenich, M.M., Leake-Jones, P.A., White, M., and Silverman, M.S. (1981) Electrode position and excitation patterns for a model cochlear prosthesis. Ann. Otol. 90, Suppl. 82.
12. Young, E.D., and Sachs, M.B. (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. J. Acoust. Soc. Amer. 66:1381.
13. Sachs, M.B., Young, E.D., and Miller, M.I. (1983) Speech encoding in the auditory nerve. Implications for cochlear implants. Ann. N.Y. Acad. Sci. (In Press).
14. Delgutte, B. (1980) Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. J. Acoust. Soc. Amer. 68:843.
15. Miller, M.I., and Sachs, M.B. (1983) Auditory-nerve representation of voice pitch. In: Hearing: Physiological Bases and Psychophysics. R. Klinke and R. Hartmann, eds. (In Press).
16. Goldstein, J.L. (1973) An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Amer. 54:1496.
17. Moore, J.C.J. (1982) An Introduction to the Phychology of Hearing, 2nd ed. Academic Press, London.
18. Evans, G.E. (1978) Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons. Audiology 17:369.
19. Kim, D.O., and Molnar, C.E. (1979) A population study of cochlear nerve fibers: Comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. J. Neurophysiol. 42:16-30.
20. Loeb, G.E., White, M., and Merzenich, M.M. (1983) A new theory of acoustic pitch perception. Biol. Cyber. (In Press).
21. White, M., Merzenich, M.M., and Loeb, G.E. (1983) Electrical stimulation of the eighth nerve in cat: Temporal properties of unit responses in the large spherical cell region of the AVCN. Ann. Otol. Rhinol. & Laryngol. (Submitted).
22. Merzenich, M.M., et al. Progress reports, NIH Contracts N01-NS-7-2367 and N01-NS-0-2337, Neuroprosthesis Contract Program, NINCDS, NIH, Washington,D.C.
23. White, M., Merzenich, M.M., and Gardi, J.N. (1983) Multichannel electrical stimulation of the auditory nerve: Channel interactions and processor design. Arch. Otolaryngol. (In Press).
24. Loeb, G.E., White, M., and Jenkins, W.M. (1983) Biophysical considerations in electrical stimulation of the auditory nervous system. Ann. N.Y. Acad. Sci. (In Press).
25. Shannon, R.V. (1983) Multichannel stimulation of the auditory nerve in man: I. Basic psychophysics. Hearing Res. (In Press).
26. Eddington, D.K., Dobelle, W.H., Mladejovsky, M.G., Brackmann, D.E., and Partin, J.L. (1978) Auditory prosthesis research with multiple channel intracochlear stimulation in man. Ann. Otol. 87, Suppl. 53:5.