*Journal of Speech and Hearing Research*, Volume 32, 93–111, March 1989

# PROSODIC AND SEGMENTAL ASPECTS OF SPEECH PERCEPTION WITH THE HOUSE/3M SINGLE-CHANNEL IMPLANT

STUART ROSEN JOHN WALLIKER
*University College London*

JUDITH A. BRIMACOMBE
*Cochlear Corporation, Englewood, CO*

BRADLY J. EDGERTON
*Medicorp Diagnostic Centers, Los Angeles*

Four adult users of the House/3M single-channel cochlear implant were tested for their ability to label question and statement intonation contours (by auditory means alone) and to identify a set of 12 intervocalic consonants (with and without lipreading). Nineteen of 20 scores obtained on the question/statement task were significantly better than chance. Simplifying the stimulating waveform so as to signal fundamental frequency alone sometimes led to an improvement in performance. In consonant identification, lipreading alone scores were always far inferior to those obtained by lipreading with the implant. Phonetic feature analyses showed that the major effect of using the implant was to increase the transmission of voicing information, although improvements in the appropriate labelling of manner distinctions were also found. Place of articulation was poorly identified from the auditory signal alone. These results are best explained by supposing that subjects can use the relatively gross temporal information found in the stimulating waveforms (periodicity, randomness and silence) in a linguistic fashion. Amplitude envelope cues are of significant, but secondary, importance. By providing information that is relatively invisible, the House/3M device can thus serve as an important aid to lipreading, even though it relies primarily on the temporal structure of the stimulating waveform. *All* implant systems, including multi-channel ones, might benefit from the appropriate exploitation of such temporal features.

Although the stated aim of many cochlear implant systems is to allow the full understanding of speech by auditory means alone, it is apparent that for most deafened people, the primary advantage of an implant will be in aiding lipreading. This is certainly so for the vast majority of people with single-channel implants, and would seem to be true even for most users of multi-channel systems, especially when they must understand speech in real-life environments with competing noise and reverberation.

For this reason, assessments of performance with implants should be heavily weighted to those features of speech that are known to be important aids to lipreading. Primary among these are the segmental and suprasegmental features associated with vocal fold activity (Fourcin et al., 1979). On the segmental level, the presence or absence of vocal fold vibration is the main cue to the phonetic feature of voicing, which differentiates, for example, "veal" (with a voiced /v/) and "feel" (with a voiceless /f/). On the suprasegmental level, variations in vocal fold vibration rate (heard as variations in voice pitch) can, for example, differentiate questions and statements, signal "new" information in an utterance, disambiguate pronoun reference, and mark syntactic units. That auditory voice pitch information can, indeed, significantly improve lipreading ability has been shown in numerous studies (e.g., Grant, Ardell, Kuhl, & Sparks, 1985; Risberg, 1974; Rosen, Fourcin, & Moore, 1980, 1981; Rosen, Moore, & Fourcin, 1979). An important part of that improvement, at least in perceiving connected speech, seems to lie in the variations of voice fundamental frequency, and not simply in its presence or absence (Rosen, Fourcin, & Moore, 1980).

Relatively little data on the perception of these features exist for the single most widely used implant, the House/3M device. This is perhaps surprising, given its long history and current use by some hundreds of people.

Two studies investigating the perception of variations in voice pitch by users of the House/3M system have appeared. Both have used the question/statement subtest of the Minimal Auditory Capabilities (MAC) battery (Owens, Kessler, Telleen, & Schubert, 1981), which requires the identification of 20 utterances as either questions or statements on the basis of their intonation contours (without lipreading). Edgerton, Prietto, and Danhauer (1983) reported a *maximum* score of 75% (chance being 50%), achieved by 4 of the 12 subjects tested. The mean score was only 65% and one half of the 12 subjects did not perform significantly better than chance. This outcome led Edgerton et al. to conclude that "... prosodic cues may not be easily discriminated by single channel cochlear implant subjects" (p. 275). Similar results have been reported by Tyler et al. (1985) on the basis of 3 subjects. Here the mean score was 58% and the maximum 80%, with 2 of the 3 subjects obtaining scores indistinguishable from chance.[1]

[1] That relatively large intonational differences *can* be perceived correctly is shown indirectly by Leder et al. (1986) in a production study involving a single subject. They report that the use of voice fundamental frequency to signal contrastive stress lost through a long period of total deafness, can be regained through use of the House/3M system.

0022-4685/89/3201-0093$01.00/0

These rather poor performances are not due to the use of a single-channel. Rosen et al. (1983) have reported scores of 100% for three subjects on the same test who were stimulated via a single extra-cochlear electrode. These subjects, however, were using the External Pattern Input (EPI) Group speech processing scheme in which only the fundamental frequency of the speech is presented. Furthermore, they were highly experienced, having participated a number of times in tests like this, and had received explicit training in the labelling of questions and statements using a visual display of voice fundamental frequency (Abberton & Fourcin, 1984).

The poor performance of users of the House/3M device may thus arise from three causes. First, the signals presented by the speech processor might be too complicated to allow the fundamental frequency to be easily perceived. Rosen and Ball (1986) already have presented evidence that a simplification of the input can lead to improved performance in identifying prosodic contours for users of the Vienna single-channel extra-cochlear implant (Hochmair & Hochmair-Desoyer, 1985). Second, the House/3M users might not have been sufficiently familiar with the task and would perform better simply through experience. The MAC test allows for no training, and typically each subtest is administered once only. Finally, a point related to the last, the subjects might not have fully understood what was expected of them. It has been frequently noted (see e.g., Rosen et al., 1985) that even normally hearing listeners can experience difficulty in making prosodic contrasts. The constraint on performance here clearly is *not* a sensory one, but one that seems to arise from the requirement to label explicitly contrasts that are only vaguely represented in our writing system (primarily through punctuation). Generally speaking, this difficulty can be avoided by appropriate training, preferably through the use of visual displays (e.g., Abberton et al., 1985).

One of the aims of the present study, then, was to determine if the ability of users of the House/3M device to identify intonation contours could be improved by signal processing, training with visual displays, or extended experience. Any improvements noted could point the way to better speech coding schemes, improved rehabilitation, or more controlled test design, respectively.

Data are sparse, too, on the use of segmental voicing information by users of the House/3M system. Testing by the group at the House Ear Institute (HEI) has tended to center on the identification of environmental sounds, performance on the MTS (monosyllable-trochee-spondee) test (which can be done on the basis of a wide variety of acoustic cues), and free-field thresholds (which give no information about the ability of subjects to distinguish among sounds).

Both Edgerton, Prietto, and Danhauer (1983) and Tyler et al. (1985) also reported scores on the MAC subtests (or adaptations thereof), which assess the identification of initial or final consonants. Although these subtests incorporate contrasts between a variety of phonetic features, the results from a subset of the test items can be used to estimate the extent to which subjects are sensitive to voicing. Edgerton, Prietto, & Danhauer (1983) found that, on average, users of the House/3M device do exhibit some ability to identify both initial and final consonants correctly, with voicing subscores better than chance. Additional indirect evidence that these subjects exhibit some ability to detect the presence or absence of voicing is found in two further findings. First, the subjects obtained nasality subscores better than chance and Rosen et al. (1985) have presented evidence that the nasality subscore may not indicate anything about the perception of nasality, per se, but simply the use of the associated voicing cue. Second, it is well known that for utterances of the type used in these tests (i.e., in isolation), vowels followed by voiced consonants tend to be longer in duration than vowels followed by voiceless consonants. The accuracy of voicing judgments with the final consonants was, in fact, significantly better than that obtained with initial consonants, and this probably reflects the subjects' sensitivity to the duration of the vowel, rather than, for example, the various dynamic spectral changes that can distinguish voiced from voiceless plosives. (See Rosen et al., 1985, for further discussions relating to the interpretation of results from the MAC battery).

The results of Tyler et al. (1985) also support the notion that users of the House/3M device are sensitive in some degree to segmental voicing information. These researchers found, with a new recording of the MAC subtests, that 2 of 3 subjects tested performed better than chance with the initial consonants, whereas all 3 achieved this level of performance with consonants in final position. Unfortunately, Tyler et al. did not do further analyses to determine which features were being used by the subjects. However, the similarity of average scores to those reported by Edgerton, Prietto, & Danhauer (at 37% and 52% for the initial and final consonants, respectively, compared to Tyler et al.'s 41% and 55%) suggests that the two sets of subjects were probably sensitive to the same speech features.

Because there had been such little investigation, and it was based on tests whose results are difficult to interpret, we aimed to examine, in a more straightforward way, the extent to which users of the House/3M device were sensitive to voicing, and whether that ability, as for prosodic contrasts, could be improved by a simplification of the signal. Furthermore, by using an appropriately constructed test, the ability to use other features of speech could be determined, in particular sensitivity to frication and the variety of acoustic cues that serve to signal place of articulation.

## METHOD

### *Subjects*

Four profoundly deaf adult cochlear implant users participated in this study. All were postlingually deafened and each had had 2 to 4 years of daily experience with the House/3M device. All were accustomed to participating in experimental investigations of their perceptual performances.

Subject 1 (S1) was a 21-year-old woman who had been profoundly deaf since age 13 due to unknown causes. It is believed that her hearing loss had been progressive since birth, and at 6 she began using a hearing aid. At age 19, she received an implant in her right ear, which had unaided acoustic thresholds of 85 dB HL at 0.25 kHz, 105 dB HL at 0.5 and 1 kHz, and no responses obtainable over the range 2 to 4 kHz. Her left ear showed similar unaided thresholds and she continued to use a hearing aid in this unimplanted ear for 1 year following implant surgery. At the time of testing, she used only the cochlear implant, no longer finding the hearing aid beneficial.

Subject 2 (S2) was a 59-year-old woman who had been profoundly deaf since age 54. Her hearing loss had been progressive since age 40 from no known cause. This subject wore hearing aids for 15 years but was recommended for cochlear implantation at the age of 55, when it was determined that amplification was providing little benefit. At the time of surgery, pure tone thresholds could not be obtained in either ear over the frequency range 0.25 to 4 kHz.

Subject 3 (S3) was a 38-year-old woman with a hereditary hearing loss that is believed to have begun in early childhood but was not diagnosed until she was 9-years-old. She began wearing hearing aids at age 9 and still wore a powerful behind-the-ear device in her unimplanted ear. Her hearing had been progressively deteriorating in the better, unimplanted ear and she was, at the time of testing, profoundly deaf bilaterally, although able to use the telephone with her acoustically aided ear. Cochlear implant surgery was performed on the poorer ear of this subject 3 years prior to testing. Preoperative unaided thresholds in the ear to be implanted were 122 dB HL at 0.5 kHz, 112 dB HL at 1 kHz, 109 dB HL at 2 kHz and 120 dB HL at 3 kHz (no tests were performed at 4 kHz and above).

Subject 4 (S4) was a 50-year-old woman who had had a progressive hearing loss since age 32 as a result of cochlear otosclerosis. She had worn a hearing aid since the diagnosis of her loss and, at the time of testing, still wore a powerful behind-the-ear hearing aid in her better, unimplanted ear. Although profoundly deaf since age 45, she too could use the telephone with her acoustically aided ear. She received the cochlear implant 3-½ years prior to testing. Preoperatively, her implanted ear showed unaided thresholds of 110 dB HL at 1 kHz, 100 dB HL at 2 kHz, and 95 dB HL at 3 and 4 kHz. No response could be obtained at frequencies of 0.25 and 0.5 kHz.

## *Device Description*

The House/3M system consists of an active electrode implanted in the scala tympani of the cochlea and a ground electrode placed in the Eustachian tube or temporalis muscle. Both electrodes are attached to an induction coil (acting as the receiver) implanted subcutaneously behind the auricle. A second induction coil (acting as the transmitter) is worn externally and aligned directly over the internal coil. A miniature, omnidirectional electret microphone senses the sound-pressure, and routes it to a signal processor where it is amplified, band-pass filtered (340–2700 Hz, with filter slopes of about 12 dB/octave) and used to amplitude modulate a 16-kHz sinusoidal carrier. The operation of this modulator is highly nonlinear, one of its most important characteristics being a hard-clipping of its output level for sound-pressure inputs greater than 65–75 dB SPL, depending upon the setting of the user-operated "sensitivity" control. (For more details on the effects this processing has on acoustic signals, see Edgerton, Doyle, Brimacombe, Danley & Fretz, 1983; Fretz & Fravel, 1985; and below). The modulator output is then amplified and transferred via electromagnetic induction across the skin to the internal coil and the active cochlear electrode. Further technical details may be found in Danley and Fretz (1982) and Fretz and Fravel (1985).

## *Test Material*

*Question/statement.* The ability to hear changes in voice fundamental frequency was tested using the question/statement subtest of the MAC battery (Owens et al., 1981) as well as new recordings based on it. In these, the same list of twenty sentences was used but the order in which they appeared as questions or statements varied. Two tests were created, spoken by a young female, native speaker of general American English, and will be referred to as the HEI-MAC.

It is well known that the ability to distinguish different frequencies of electrical stimulation applied to a single channel decreases as frequency rises from 100 to 500 Hz. By the use of both a male and a female speaker, we hoped to investigate the possibility that differences in the average fundamental frequency of the speaker could lead to differences in the ability of subjects to perceive prosodic contours (Rosen et al., 1985). Measurements made from the recorded question/statement material show the male speaker in the original MAC recording to have a mean fundamental frequency—on a logarithmic scale—of 98 Hz (with 90% of his voice-pitch periods between 66 and 163 Hz) whereas our female speaker exhibits a mean fundamental frequency of 231 Hz (with 90% of her voice-pitch periods between 164 and 365 Hz).

*Intervocalic consonants.* The segmental test used a set of 12 vowel-consonant-vowel (VCV) utterances in which the vowel was always /ɑ/ ("ah") and the consonants were /m/, /b/, /p/, /v/, /f/, /d/, /n/, /z/, /s/, /t/, /g/ or /k/ (Rosen et al., 1979). This particular set of consonants was chosen to sample a variety of distinctions in voicing, place of articulation, and manner of articulation.

A test consisted of four repetitions of each of the 12 VCVs in a random order subject to the constraint that each consonant occurred twice in the first 24 trials. Two such random orders were created and were spoken by the same young woman who recorded the question/statement test. All stimuli were uttered with primary stress on the first syllable, and a falling intonation.

General American speech patterns in such a situation would ordinarily lead to the /t/ being flapped, losing its aspiration entirely and becoming more /d/-like (as in "whiter" typically being realized as something closer to "wider"). The speaker was instructed explicitly to aspirate the /t/, and neither she, nor the subjects, seemed to find this odd.[2]

*Recording procedure.* All the tests (except the original MAC test) were recorded in a single session using a Sony U-Matic video recorder and color camera. The speech signal from a microphone was recorded on one audio channel, and on the other, a series of pulses whose fundamental frequency followed, in real time, the voice fundamental frequency of the speaker. These were derived from a temporally based microphone-driven fundamental frequency extractor (Howard & Fourcin, 1983).

## *Conditions of Presentation*

*Auditory stimuli.* To determine if simplifying the speech signal had any advantages over the current House/3M system, we explored the possibility of using two types of signal to transmit voice pitch information to the implanted subjects while bypassing their normal speech processor—triangular pulses, and bursts of a 16-kHz sinusoidal carrier.

Triangular pulses were of particular interest because they result in a simple waveform at the internal electrode. Due to the differentiation of the signal effected by its transmission between the transmitting and receiving coil, 2-ms triangles became biphasic pulses of 1 ms per phase, with the negative phase leading at the active electrode. This was confirmed by examining the voltage arising at the stimulating electrode of a "dummy receiver," a receiving coil and resistor meant to simulate the in-vivo impedance at the intrascalar electrode (however imperfectly).

Two properties of these triangular pulses, not previously used with the House/3M system, needed to be explored. First, there had to be a reasonable dynamic range across the frequencies of interest (about 100 to 400 Hz, the main area for adult voice fundamental frequencies), so that we could stimulate with a constant pulse amplitude without causing an uncomfortable loudness. As measured at the input to the transmitting coil, S1 had a dynamic range of about 7 dB (i.e., a voltage ratio of 2.24: 1) at 400 Hz, and a little larger than this at 100 Hz. S3 had dynamic ranges of 9 dB at 400 Hz and 14 dB at 100 Hz. These were more than adequate for our purposes. (S2 and S4 were not available for these tests).

The second property investigated was the way in which perceived loudness varied with frequency. This was to ensure that performance on the question/statement task was indeed a result of perceiving subjective pitch changes and not associated loudness changes. For S3, equal amplitude trains of triangular pulses at 100, 200, and 400 Hz were perceived as being close in loudness. For S1, equal loudness was attained when the 400 Hz triangles were −4 dB, and the 200 Hz triangles were −3 to −2 dB, re the amplitude of the 100 Hz triangles. That this degree of loudness change with stimulating frequency was acceptable for testing purposes is supported by the fact that S1 could not detect a 2-dB change in voltage when both stimuli were at 100 Hz. (Again, the other two subjects were not available for testing).

Because the House/3M device normally works (for more efficient power transmission) with a 16-kHz carrier that is modulated by a relatively low-frequency signal, we also used 1-ms bursts of a 16-kHz carrier triggered at the appropriate rate. Only S3 was available for psychophysical testing with this signal, and she had a 7–8 dB dynamic range at both 100 Hz and 400 Hz. A 400 Hz signal whose amplitude was −2 dB re that of 100 Hz bursts led to equal loudness of the two signals.

The signal provided most frequently was speech. Most testing was done with a direct connection between the video- or tape-recorder and the subjects' speech processor, avoiding the microphone link and thus ensuring better stimulus control and the prevention of any interference from background noise. Some testing was also done free-field, with subjects sitting in an IAC sound-treated room about 1 m from a loudspeaker. Stimuli were presented at a level of about 70 dB SPL at the microphone of the wearable speech processor. For all tests with speech, subjects set the sensitivity and volume controls of their speech processors to the settings that they desired.[3] When being tested by direct electrical connection, the amplitude level of the speech input to the processor was set to elicit waveforms from it that were typical of those generated in face-to-face free-field conditions. For 3 subjects, this setting produced an output that was heavily clipped whereas for 1 subject (S4), the output waveform was mainly unclipped.

*Visual stimuli.* For the assessment of visual and audiovisual performance in identifying intervocalic consonants, the face of the speaker was displayed for the subject to lipread on a color video monitor.

## *Procedure*

Testing in the question/statement task always began with presentation of the original MAC subtest via direct electrical connection to the speech processor. The order of presentation of other conditions varied depending on the subject's performance and their availability for testing. All testing was done over the course of 3 days. All

[2] In more recent recordings, the speaker has been instructed to place primary stress on the second syllable of the VCV, thus avoiding the difficulty of the flapped /t/.

[3] Subjects typically used the middle sensitivity setting of the three possible. In free-field testing, however, S1 used the highest sensitivity and S4 sometimes used the lowest sensitivity.

subjects except S1 finished the testing in a single session. No training was given prior to the testing, although some of the subjects had previously participated in similar experiments using the original MAC tapes.

There were five conditions under which the VCV stimuli were presented:

1. LIPS + SP:DT—lipreading with speech input by direct electrical connection to the speech processor of the cochlear implant.
2. LIPS + SP:FF—lipreading with speech input presented in a free-field.
3. LIPS—lipreading alone.
4. SP:DT—speech alone presented by direct electrical connection to the speech processor.
5. SP:FF—speech alone presented in a free-field.

The order in which the different testing conditions occurred was quasi-random. The number of tests in each condition by each subject was limited by time, and by the desire both to investigate the range of possible conditions, and to collect a fairly large number of responses in the condition that seemed most important to us (LIPS + SP:DT). There was a total of 53 sessions resulting in 2544 separate responses.

## RESULTS AND DISCUSSION

### *Question/Statement*

Table 1 presents the results from the question/statement tests. For each subject, the scores are presented in the order in which the tests were performed. Although some care in interpretation is required given the limited amount of testing (still large compared to typical implant studies), the data reveal a number of interesting findings.

TABLE 1. Results from the question/statement test.

| *Subject* | *Test* | *Condition* | *Score (% of 20)* |
|---|---|---|---|
| S1 | MAC | SP:DT | 90 |
| | HEI-MAC | TR:DT | 70 |
| | HEI-MAC | 16k:DT | 90 |
| | HEI-MAC | SP:DT | 90 |
| | HEI-MAC | SP:DT | 75 |
| | HEI-MAC | 16k:DT | 100 |
| | HEI-MAC | SP:FF | 90 |
| | MAC | SP:FF | 85 |
| | MAC | SP:DT | 85 |
| S2 | MAC | SP:DT | 75 |
| | HEI-MAC | SP:DT | 95 |
| S3 | MAC | SP:DT | 75 |
| | HEI-MAC | SP:DT | 100 |
| | MAC | TR:DT | 100 |
| | HEI-MAC | TR:DT | 95 |
| | MAC | 16k:DT | 100 |
| | HEI-MAC | 16k:DT | 90 |
| | HEI-MAC | SP:DT | 90 |
| S4 | MAC | SP:DT | 100 |
| | HEI-MAC | SP:DT | 65 |

*Note.* Under "Condition," the two symbols before the colon refer to the type of waveform used: SP(eech), TR(iangles) or 16k(Hz carrier in 1-ms bursts). The two letters after the colon indicate whether the stimuli were presented by *D*irec*T* connection to the speech processor, or in a *F*ree *F*ield. Scores of 70% or better are statistically significantly different from chance at the .0577 level (exact binomial probability).

First, all subjects performed at significantly better than chance level on the original MAC test (male speaker), with scores between 75% and 100% (mean of 85%). In fact, of the 20 tests performed, only one score failed to reach statistical significance (S4 listening to the speech signal of HEI-MAC). This level of performance is considerably better than that reported for users of the House/3M device by either Edgerton, Prietto, and Danhauer (1983) or Tyler et al. (1985). Either these subjects were intrinsically better performers than is typically found, or their more extensive experience in being tested allowed them to overcome the problems that subjects often have in this task. Most likely, both factors were operating. Of Tyler et al.'s 3 subjects, one of those who performed poorly had been implanted for only 6 months. In any case, he wore his device less than the other 2 subjects, relying more on a hearing aid in the unimplanted ear. Edgerton, Prietto, & Danhauer 's subjects seem more comparable, at least in terms of implant use, to those in the present study. All had had at least 1 year of daily experience with their implants.

Secondly, on average, subjects seemed to perform equally well with both speakers. Although S2 and S3 achieved higher scores when the speaker was a woman, this might reflect an order effect as they heard the male speaker first. Interestingly, S4 did considerably worse with the female speaker, and this may well reflect extra difficulties with a higher-pitched voice.

Thirdly, for the 1 subject for which data were collected via both free-field and direct connection, there was little difference between the scores obtained.

Finally, both simple waveforms (triangles and 1-ms bursts of 16-kHz carrier) resulted in high levels of performance. This is especially striking for S3, who scored perfectly the first time she was tested with the triangles, even though this was a waveform totally unlike any she receives in her daily experience with the implant. Furthermore, there is some evidence that such simplified signals lead to better percepts of voice pitch contours than those obtained with speech. S1's mean performance on the HEI-MAC via direct connection was better with 1-ms bursts of 16-kHz carrier (95%) than with speech (82.5%), a difference that is significant at the .05 level (one-tailed test of binomial proportions). In contrast, S3 showed essentially no differences with changes in stimulating waveform (ignoring the first session, whose relatively low score presumably arises from relative inexperience), with all scores in the 90% to 100% range. However, because her performance was so high with the speech signals, it is difficult to demonstrate any improvement when other waveforms are employed.

Further tests of this sort should employ stimuli that require more acute discrimination (for example, the synthesized continua of Rosen & Ball, 1986), both to avoid ceiling effects and because there is useful information cued by changes in fundamental frequency that are smaller than those displayed in citation question/statement forms. The intonation contours in the current question/statement tests typically traverse a wide range (on the order of an octave), and may be considered only a simple test of subjects' abilities to use fundamental frequency changes in speech.

### *Intervocalic Consonants*

As a preliminary to all further analyses, confusion matrices were constructed for each individual testing session, and summed within subjects to obtain 20 mean matrices (5 conditions × 4 subjects). From these, a number of summary statistics were taken (Table 2), most aimed at determining the extent to which a particular phonetic feature was being perceived by the subject. The three major articulatory/phonetic categories were used: *place* (bilabial /mbp/ vs. labiodental /vf/ vs. alveolar /nzdts/ vs. velar /gk/)[4]; *voicing* (voiced /mnvzbdg/ vs. voiceless /ptkfs/); and, *manner* (plosive /bdgptk/ vs. fricative /vfzs/ vs. nasal /mn/). An information transfer measure (Miller & Nicely, 1955) was calculated for each feature, primarily because this permits a direct comparison of performance among features that vary in the number of possible categories, and the varying number of phonemes within each category. Because the ordering of the scores for the different features appears similar across subjects within a particular condition, we also summed results over subjects to obtain 5 matrices that represented all the data obtained for each condition. The results of information transfer analyses of these matrices are presented graphically in Figure 1.

Compare first the results obtained from lipreading alone and lipreading with the implant via direct connection. (As there were no significant differences in performance on any measure between the summed LIPS + SP: DT and LIPS + SP:FF matrices, only performance in the more frequently tested direct connection condition will be discussed).

Figure 2 gives an overview of mean subject performance in the form of a graphical confusion matrix. When performing the task by lipreading alone, responses fell neatly into three or four classes determined by place of articulation. As can be seen in both Figures 1 and 2, the addition of auditory information through the cochlear implant greatly increased overall performance relative to that obtained by lipreading alone.

TABLE 2. Summary statistics for each subject in the intervocalic consonant test presented under 5 conditions.

| | *Subject* | | | |
|---|---|---|---|---|
| *Condition* | *S1* | *S2* | *S3* | *S4* |
| LIPS+SP:DT | | | | |
| % correct | 95 | 88 | 87 | 79 |
| % information | | | | |
| place | 97 | 80 | 88 | 86 |
| voicing | 80 | 89 | 70 | 58 |
| manner | 97 | 82 | 81 | 93 |
| no. of sessions | 4 | 6 | 4 | 4 |
| LIPS+SP:FF | | | | |
| % correct | 97 | 90 | 76 | 84 |
| % information | | | | |
| place | 96 | 85 | 83 | 91 |
| voicing | 92 | 85 | 48 | 70 |
| manner | 95 | 91 | 83 | 87 |
| no. of sessions | 2 | 2 | 2 | 2 |
| LIPS | | | | |
| % correct | 40 | 44 | 43 | 44 |
| % information | | | | |
| place | 91 | 87 | 78 | 83 |
| voicing | 2 | 3 | 4 | 3 |
| manner | 39 | 29 | 40 | 42 |
| no. of sessions | 3 | 2 | 3 | 2 |
| SP:DT | | | | |
| % correct | 42 | 36 | 34 | 37 |
| % information | | | | |
| place | 11 | 7 | 9 | 17 |
| voicing | 66 | 66 | 39 | 51 |
| manner | 70 | 43 | 52 | 48 |
| no. of sessions | 3 | 3 | 3 | 3 |
| SP:FF | | | | |
| % correct | 35 | 27 | 23 | 27 |
| % information | | | | |
| place | 12 | 13 | 6 | 18 |
| voicing | 67 | 87 | 51 | 46 |
| manner | 22 | 38 | 27 | 31 |
| no. of sessions | 2 | 1 | 1 | 1 |

Of the three phonetic features, the ability to make voicing contrasts was improved most by the addition of electro-auditory information to lipreading. Subjects received very little information about voicing when they were only lipreading, reflecting the fact that the primary articulatory correlate of voicing (presence or absence of vocal fold vibration) is invisible. The high proportion of information transfer about voicing when the implant was combined with lipreading (corresponding to error rates of only 5%) shows that the device provided strong cues about the primary acoustic correlate of voicing (presence or absence of low-fundamental-frequency quasi-periodic sound).

Perception of manner, too, was improved with the implant, although some manner information is available from visual cues. Part of the visible manner information arises from the interdependence of the phonetic features. For example, once a consonant is seen to be a labiodental (where errors are almost never made), the subject knows it must be a fricative, and not a plosive or a nasal. By the

[4] Other studies have defined the place contrast differently. Dowell et al. (1982) used a place feature that does not distinguish between bilabial and labiodental (lumping them both together into a category "front"), simply because that was the definition used by Miller and Nicely (1955). Such a classification may be sensible for the early experiments, as they were based on the identification of consonants by sound alone. In a study that uses lipreading, the bilabial/labiodental distinction is one of the simplest to make, and to define a feature that ignores one of the strongest aspects of the data seems slightly perverse. For a lipreading study using the present set of consonants, the only alternative definition of the place categories that might be useful is one that treats all the back consonants (alveolars and velars) as one class, because these are difficult to distinguish visually.

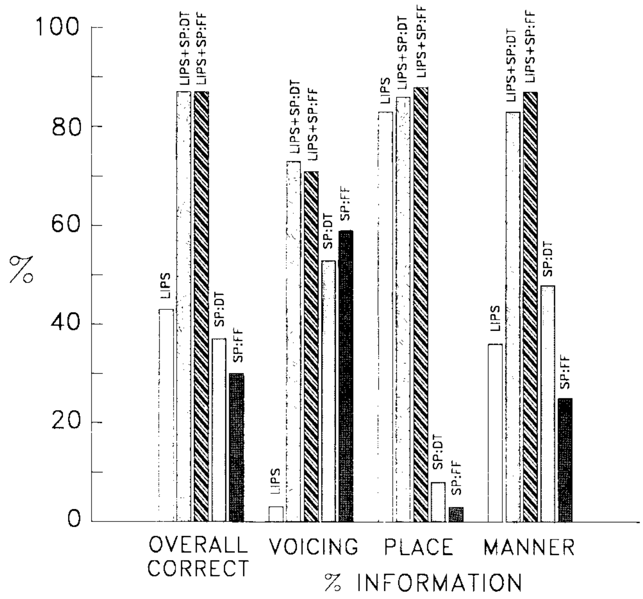

FIGURE 1. Statistics derived from confusion matrices summed over all subjects. The first set of bars depicts overall performance in terms of percent correct, whereas the others depict performance for each of the three traditional articulatory/phonetic dimensions, calculated with an information transfer statistic.

same token, there are no bilabial fricatives in this set of consonants. There do, however, also seem to be visual cues about manner per se, presumably contained in small differences in the duration and details of articulatory maneuvers. An information transfer analysis of submatrices of the mean confusion matrix obtained by lipreading alone shows that even within the alveolar sounds, 23% of the manner information is successfully received. By contrast, only 4% was received within the bilabials.

That some degree of manner information independent of that related to place of articulation may be transmitted by visual cues alone is also supported by a Sequential Information Analysis (SINFA, Wang & Bilger, 1973). SINFA is a way to partial out the effects of relationships and redundancies among a set of prescribed features. It begins by identifying the single feature that leads to the highest proportion of information transfer, and then calculates the importance of the remaining features using a measure conditional on the performance according to the first chosen feature. Again, the most important feature is chosen, and the information transfer of the remaining features calculated conditional upon that obtained for the first two chosen features. By iterating this procedure, one thus obtains an estimate of the importance of each (now independent) feature, in terms of the proportion of total information transferred that each of them can account for. SINFA applied to the overall matrix obtained by lipreading alone shows that 8% of the transmitted information was due to manner distinctions, even after the role of place (which accounted for 81%) had been partialled out. Voicing only accounted for 1% of the information successfully transferred.

Just as it appears that some ability to make manner distinctions may be conditional upon distinctions in place of articulation, so too may part of the improvement in perception of manner features with use of the implant be attributed to sensitivity to voicing. Some visually confusable consonants (i.e., those at the same place of articulation) differ from one another in both manner and voicing (e.g., /m/ and /p/). Therefore, even if the subjects are sensitive only to voicing, they may improve their perception of manner as well. SINFA of the summed LIPS + SP: DT matrix shows though, that some 17% of the transmitted information is manner-related even after the place and voicing features are partialled out.[5]

However, even "pure" manner distinctions can be made on the basis of the acoustical reflection of voicing activity. For example, the nasal /m/ is continuously voiced, whereas the voiced plosive /b/ may have a short gap in its voicing. Even if voicing is present, the acoustic energy during the closure tends to be very low, certainly much lower than during the closure for /m/. This distinction in the *acoustic* details of voicing activity is not incorporated in the *phonetic* feature of voicing, where both /m/ and /b/ are classified as "voiced." That subjects are perceiving more auditory manner information than can be accounted for by the acoustic details of voicing activity only is supported weakly by a SINFA with a three-valued "voicing" feature (distinguishing voiceless /fsptk/ from voiced sounds with low-amplitude periodicity during the consonantal closure /vzbdg/ from voiced sounds with high-amplitude periodicity /mn/). Manner accounted for about 7% of the total information transmitted in both LIPS + SP conditions after the place and three-valued voicing information were partialled out. Clearer evidence that subjects *were* sensitive to more than just voicing information is found in the results obtained without lipreading, discussed below, where there is evidence that subjects are able to perceive voiceless excitation, and distinguish it from silence and voicing.

Finally, the place feature was well perceived in all three lipreading conditions, not surprisingly given the strong visual cues to place. Furthermore, the near equality of the value of the information transfer statistic for conditions with and without the implant lends at least weak evidence to the supposition that the acoustic cues to place of articulation are not being used.

Much stronger evidence for this was found when subjects were required to identify the sounds without visual

[5] SINFA may also be used in a mode in which the order of evaluation for the different features is predetermined, rather than based on the relative efficiency of transmission. In its normal mode of opertion on the LIPS + SP:DT matrix, SINFA identified place, manner, and voicing in that order. However, the efficiency of transmission for manner and voicing differed by very little after the effect of place was partialled out. Therefore, because we were specifically interested in the degree of manner information over and above that associated with voicing, SINFA was instructed to analyze place and voicing first. This option, of determining the order of feature identification, has also been used in a number of other SINFA calculations cited.

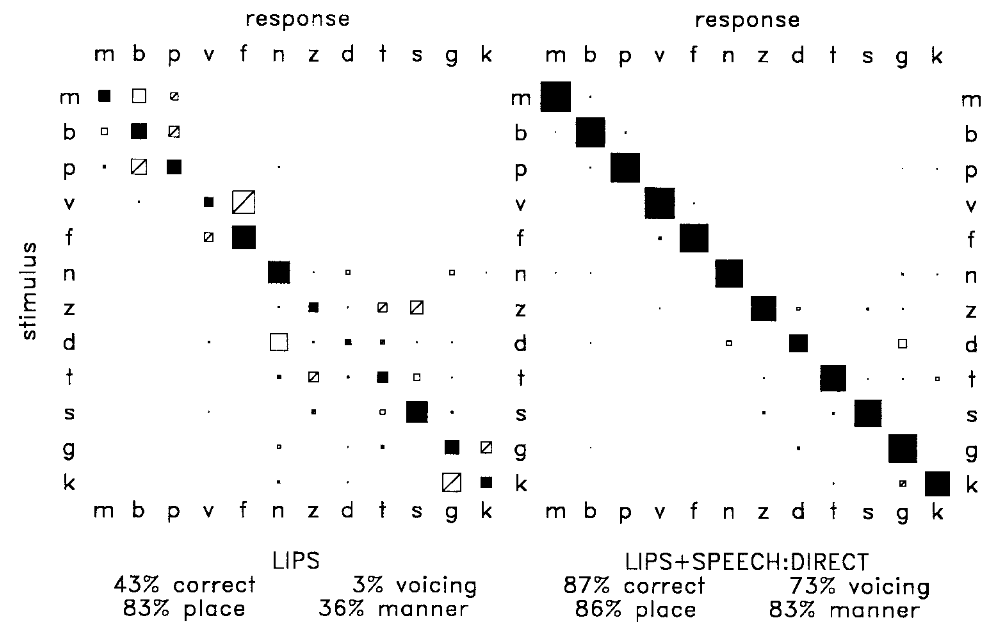

FIGURE 2. Graphical confusion matrices, summed over all subjects, comparing performance by lipreading alone (LIPS) and by lipreading with speech input by direct connection (LIPS + SPEECH:DIRECT). The length of the side of each square is proportional to the number of times that a particular stimulus (given by the row) was identified as the column response. Correct responses (down the diagonal) are solid black, whereas voicing errors are indicated by a diagonal line through the square. The overall percent correct and percent of information transferred for voicing, place, and manner is indicated under each matrix.

cues. Figure 3 shows the mean confusion matrices obtained, for direct and free-field presentations, again in a graphical form. The order of the stimuli and responses has been changed to emphasize the structure in the data, the new order determined by the manner and voicing characteristics of the sounds. These go from, in some sense, the "most voiced" sounds (the nasals) to the "least voiced" (the voiceless fricatives). Clearly, it is these two

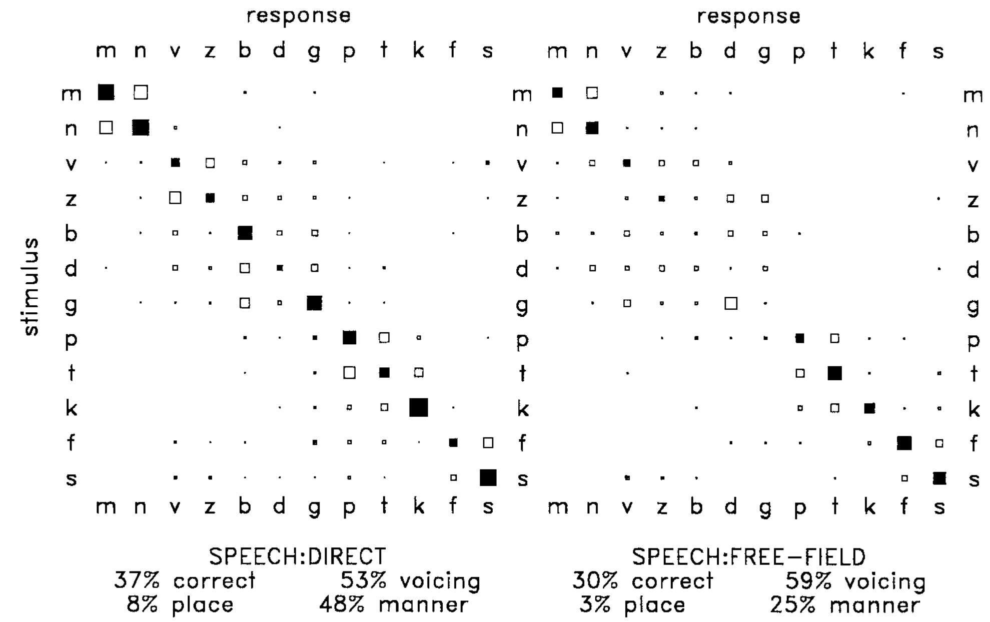

FIGURE 3. Graphical confusion matrices, summed over all subjects, comparing performance by implant alone via direct connection (SPEECH:DIRECT) and via free-field presentation (SPEECH:FREE-FIELD). Format as for Figure 2, save in that the order of the stimuli and responses has been changed to group together sounds with the same manner and voicing, rather than the same place of articulation.

features that were the most important determinant of the confusions that occurred. A five-valued manner/voicing feature (consisting of voiced nasals, voiced fricatives, voiced plosives, voiceless plosives and voiceless fricatives) accounts for 87% of the information transferred in SP:DT, and 75% in SP:FF.

That place was fairly inconsequential is shown also in the feature analyses (Figure 1). Whatever sensitivity to place *was* exhibited may have resulted from reliance on what are voicing or manner features. This again may be based on interdependencies among the defined features: continuously voiced nasals can only be bilabial or alveolar; fricatives can only be labiodental or alveolar. The results of SINFA however, revealed that even after the contribution of a five-valued voicing/manner feature is partialled out, place accounts for 6% and 13% of the total information transmitted in conditions SP:DT and SP:FF, respectively. Note, though, that this finding does *not* imply that subjects were sensitive to the primary acoustic cues to place of articulation—the static and dynamic distribution of energy across frequency. First, it may be that contributions of 6% to 13% of total transmitted information arise from defining only a small number of features, and/or are not significantly different from chance. (SINFA uses no inferential statistics and has an arbitrary criterion for deciding whether or not a feature is important in determining the pattern of subject errors.) Second, subjects may have been using information about the place of articulation contained in the details of voicing activity. For example, voice onset time in plosives is known to covary in an orderly way with place (Lisker & Abramson, 1964).

An empirical way to assess the relevance of such transmission of "place" distinctions is to perform experiments in which only certain acoustic information is present, and then to assess the "conclusions" of SINFA. Two normal listeners, experienced in participating in such experiments, attempted to identify this set of intervocalic consonants on the basis of fundamental frequency patterns alone. The auditory signal was generated by triggering narrow rectangular pulses from the recorded pitch-extracted pulses (see under Recording Procedure). Two sessions were run, one of each of the two sessions recorded, and the resulting data summed into one confusion matrix. SINFA showed the transmission of "place" information to account for 9% of the total information transferred after partialling out the effects of a five-valued manner/voicing feature. As the fundamental frequency information did not contain any of the primary acoustic cues to place of articulation (i.e., relating to spectral variations), the degree of transmission of "place" information evidenced by users of the House/3M device is not large enough to infer that they are sensitive to spectral cues.

In conditions performed by audition alone (SP:DT and SP:FF), voicing was the best perceived of the three phonetic features. Although there is little or no visual voicing information (as determined in LIPS), the voicing information perceived by the subjects was considerably higher when they were allowed to lipread than when they were only listening. It may be that the visual information cues the arrival of a sound, or makes the subject more confident. Or, again, there are feature dependencies to consider. Of those sounds that are easily identified visually, for example, ⅔ of the bilabials, but only ½ of the labiodentals, are voiced.

The transmission of manner information is particularly interesting for two reasons. First, although the overall score for free-field and direct presentations was little different (although consistently better in DT for all 4 subjects), the perception of manner was considerably worse in the free-field condition, by about a factor of ½. (The perception of place, too, decreased by one-half, but was poorly transmitted even for a direct connection.) Inspection of the appropriate confusion matrices collapsed across place into the five manner/voicing subgroups shows no single cause of this difference in the patterns of performance (Table 3). Among the voiceless phonemes, there was little difference in overall performance between the two conditions, except that both the plosives and the fricatives were more frequently labelled as fricatives in the free-field condition. Perhaps voiceless excitation was more strongly represented in the free-field. Among the voiced phonemes, errors in manner for the nasals and plosives were, more or less, uniformly more frequent in the free-field condition. However, whereas voiced fricatives and plosives were distinguished reasonably well from one another when presented directly, they led to more or less identical responses under a free-field presentation. Furthermore, voiced fricatives were most often labelled as voiced plosives under free-field conditions whereas they were most frequently labelled correctly when presented directly. Thus, it may be that the fricative information in the voiced fricatives is lost in free-field conditions. This would not, however, account for the greater number of errors with voiced plosives, nor would it be consistent with the greater number of fricative judgments for the voiceless sounds. This issue will be discussed again below in relation to the waveforms generated by the implant processor.

Secondly, as alluded to previously, there is strong evidence that these subjects can be sensitive to more than voicing information. For one thing, they showed a slight increase in their use of voicing information at the same time as there was a decrease in their abilities to make manner distinctions in going from direct to free-field conditions. More importantly, subjects were able to distinguish the presence or absence of voiceless frication, as evidenced in the collapsed confusion matrices of Table 3. If subjects were only sensitive to voicing, then they would not be able to distinguish the voiceless plosives /ptk/ from the voiceless fricatives /fs/ because all of these sounds are characterized by a relatively long period of voicelessness.

That sensitivity to voicing alone would not allow this distinction to be maintained is clearly seen in the results of the previously mentioned experiment in which normal listeners identified the intervocalic consonants on the basis of fundamental frequency patterns alone. Table 4 compares the collapsed confusion matrix obtained from the higher

TABLE 3. Mean collapsed confusion matrices for conditions which were performed by auditory means alone.

| | | *SP:DT Responses* | | | | | *SP:FF Responses* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *mn* | *vz* | *bdg* | *ptk* | *fs* | *mn* | *vz* | *bdg* | *ptk* | *fs* | | |
| Stimuli | mn | 90 | 4 | 6 | | | 80 | 10 | 8 | | 3 | mn | Stimuli |
| | vz | 4 | 56 | 30 | 2 | 7 | 15 | 38 | 45 | | 3 | vz | |
| | bdg | 2 | 17 | 75 | 5 | 1 | 17 | 33 | 47 | 2 | 2 | bdg | |
| | ptk | | | 12 | 87 | 2 | | 3 | 10 | 77 | 10 | ptk | |
| | fs | | 13 | 11 | 17 | 59 | | 13 | 8 | 10 | 70 | fs | |
| | | 215/576 correct = 37%<br>% information<br>Voicing = 53% Place = 8% Manner = 48% | | | | | 71/240 correct = 30%<br>% information<br>Voicing = 59% Place = 3% Manner = 25% | | | | | | |

*Note.* Each cell entry indicates the percentage of times that a response in the set designated by the column was given to the stimuli in the set designated by the row. Values are rounded to the nearest percent. The summary statistics under each matrix (% correct, and information transfer measures) are calculated from the original 12 × 12 matrices.

scoring normal listener with that obtained from S1. Note that the presentation of the fundamental alone did not permit the normal listener to distinguish voiceless plosives from voiceless fricatives. S1 clearly discriminated between these two classes. Similarly, her accuracy in identifying voiced fricatives was considerably better than that obtained by the normal listener on the basis of the fundamental alone, where they are not distinguished from voiced plosives. Further evidence of S1's ability to make use of frication is found in the confusions made between voiced and voiceless fricatives, confusions never made on the basis of the fundamental alone. Finally, SINFA was applied to the two original confusion matrices shown collapsed in Table 4, with three binary manner dimensions (nasality, plosion, and frication) substituted for the three-valued manner feature used so far (in addition to the previously defined features of voicing and place). For the normal listener, only the features of voicing, nasality, and place were identified as important. Therefore, according to the criterion specified for the analysis, frication must have accounted for less than 1% of the total information transmitted. For S1, all four of the above-mentioned features were identified as important (only plosion was not identified), with frication accounting for nearly 26% of the total information transmitted. Similar results were obtained from the same analysis of the SP:DT matrix summed over all subjects.

In short, the presentation of fundamental frequency alone allows the separation of the stimuli into three classes, whereas the House/3M device allows subjects to distinguish five. This is why the transmission of manner information is so much higher for S2 than it is for the normal listener operating on the basis of fundamental frequency information only.

This result also provides a preliminary answer as to whether a simplification of the signal to fundamental frequency alone (as advocated by Fourcin et al., 1979) would be desirable. Because the implant users, with their complex waveforms, can be more accurate at identifying

TABLE 4. Collapsed confusion matrices for one normal listener listening to fundamental frequency information alone, and the patient who scored best in the condition SP:DT wearing the House/3M device.

| | | *Normal listener Responses* | | | | | *House/3M - S1 Responses* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *mn* | *vz* | *bdg* | *ptk* | *fs* | *mn* | *vz* | *bdg* | *ptk* | *fs* | | |
| Stimuli | mn | 94 | | 6 | | | 96 | | 4 | | | mn | Stimuli |
| | vz | | 25 | 75 | | | | 88 | | | 12 | vz | |
| | bdg | 4 | 21 | 75 | | | | 36 | 61 | 3 | | bdg | |
| | ptk | | | 4 | 88 | 8 | | | 6 | 92 | 3 | ptk | |
| | fs | | | | 94 | 6 | | 12 | | 4 | 83 | fs | |
| | | 24/96 correct = 25%<br>% information<br>Voicing = 92% Place = 7% Manner = 38% | | | | | 61/144 correct = 42%<br>% information<br>Voicing = 66% Place = 11% Manner = 71% | | | | | | |

*Note.* Table format as for Table 3.

VCVs without lipreading than a highly practiced normal listener with fundamental frequency alone (as seen in Table 4), it seems highly unlikely that their performance will be improved by simplification. In fact, performance would almost certainly be reduced with fundamental frequency information only (at least in this particular task), because the subjects currently show sensitivity to the presence or absence of frication.[6]

Unfortunately, there was insufficient time to permit testing of the subjects with simplified signals, except for one instance. S3 was tested in one session of lipreading plus voice fundamental frequency signalled by 1-ms pulses of a 16-kHz carrier (triggered by the recorded pulses derived from the acoustic pitch extractor) via direct connection (denoted LIPS + Fundamental:DT). Her overall performance (in percent correct) under this condition was slightly worse (at 71%) than that obtained in LIPS + SP:DT (87%) and LIPS + SP:FF (76%). The unconditional transmission of manner, and specifically frication, information was considerably better in the two LIPS + SP conditions than in LIPS + Fundamental:DT.

In summary, it seems that these subjects would not benefit from an extreme simplification of the signal, and that they are better served by their current device than they would be with one that signalled fundamental frequency only (at least for the task of identifying VCVs).

## CORRELATIONS BETWEEN PATIENT PERFORMANCE AND THE OUTPUT WAVEFORMS OF THE SPEECH PROCESSOR

In our attempts thus far to account for the perceptual results obtained, we have used *a priori* knowledge about the acoustic structure of speech sounds to make informed guesses about the signals that the implant must be presenting to the patients. However, the complicated and nonlinear nature of the speech processor makes it difficult to predict its output for the set of speech sounds we used. We therefore made direct measurements of the output of the speech processor to the stimuli employed (VCVs and question/statement utterances), in both free-field and direct conditions using the dummy receiver. Most of the measurements were made with S2's processor, which functioned in a manner similar to the processors of S1 and S3. A smaller number of measurements were made with S4's processor, whose outputs were much less heavily clipped than those of the other three.

Stimuli were presented to the speech processors in conditions identical to those used in testing. The output of the dummy receiver was recorded on one channel of an audio-tape running at 15 inches per second. On the other channel the speech signal presented to the processor was recorded, either "direct" from the video recorder, or from a microphone mounted next to the ear-level mike of the patient's speech processor (free-field condition). Signals were digitized for further analysis by playing them at half speed, and sampling with 12-bit resolution at a frequency of 50 kHz.[7] No low-pass anti-aliasing filter was used because there was little energy in the signals above 20 kHz, and we wanted to avoid the phase distortion such filters typically introduce. Note that all the processor output waveforms consisted of a modulated 16-kHz carrier, but that the time scale of the oscillograms presented here is not sufficiently expanded to resolve single cycles of the carrier.

### *Intervocalic Consonants*

Figure 4 shows the output of S2's processor to each of the first 12 VCVs of Session 1, when directly connected, along with the associated speech waveforms. (Examination of the second occurrence of each stimulus revealed that they were similar to those in Figure 4.) It appears from these plots that, as long as subjects are sensitive to aspects of the amplitude envelope, they will have some success in classifying each stimulus as belonging to one of three manner/voicing classes on the basis of auditory information alone: (a) voiceless plosives, (b) nasals, and (c) fricatives and voiced plosives.

For the identification of voiceless plosives, there are two main cues—the release burst (which was accentuated by the compressive nature of the processor) and the silent gap between the first vowel and the release. Note that the processor gave a nonzero output even in these silent gaps, as well as during the silence preceding the utterance (Fretz & Fravel, 1985). The voiced nasals were distinguished from all other sounds by the presence of a constant high-amplitude excitation, although the output waveform corresponding to the nasal murmur was not clipped to the same extent as it was during the vocalic portions. Both voiced fricatives and voiced plosives showed a relatively low amplitude signal during the consonantal gesture, and the similarity of their output oscillograms was reflected in the patients' relatively frequent confusions between these two classes.

If subjects were only sensitive to amplitude envelope, however, we would expect voiceless fricatives to be put into this group as well. In fact, although there were some confusions like this, on average all 4 subjects showed some ability to distinguish voiceless fricatives from all

[6] In fact, Fourcin et al. (1979) have already proposed to signal the presence of voiceless frication with an appropriately aperiodic signal, in addition to the periodic fundamental frequency information. That subjects with single-channel extra-cochlear electrodes will be able to use this information is supported by psychophysical studies that show good performances in a task that requires the discrimination of periodic sounds (sinusoids) from aperiodic ones (bands of noise).

[7] Waveforms of the recorded stimuli were compared at both playback speeds at low frequencies (in the original speech) and high (in the speech processor output) in order to eliminate the possibility of any distortions imposed by the change in speed. In all cases, apart from the time axis distortion, the waveforms were essentially identical.

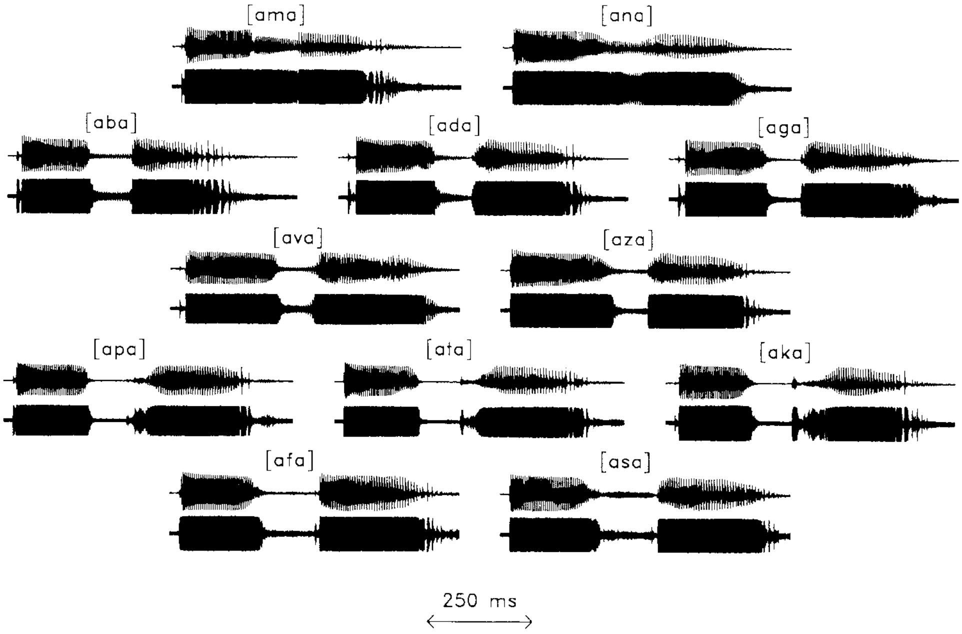

FIGURE 4. Input to the speech processor and its output for each of the twelve intervocalic consonants, via direct connection. For each utterance, the upper waveform of the pair is the speech-pressure waveform, while the lower is the output of S2's speech processor, as recorded with a dummy receiver.

other sounds (although there is a wide range of performance). This was almost certainly due to the patients being sensitive to whether or not the excitation was quasi-periodic (from vocal fold activity) or aperiodic (from turbulence noise generated by articulators that are close together in the vocal tract).

A closer look at the waveforms generated by the speech processor shows that they did, indeed, reflect the degree

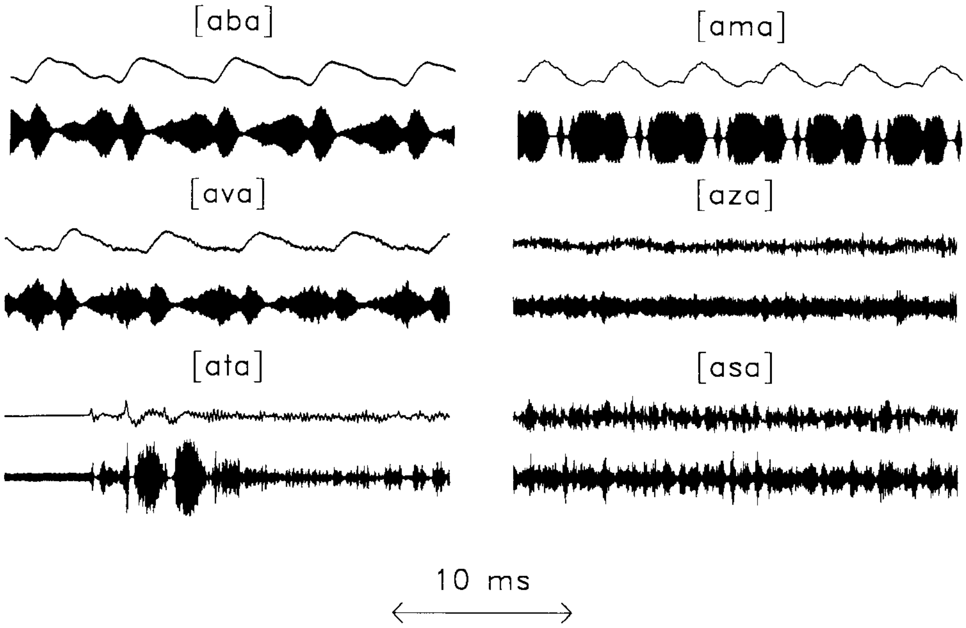

FIGURE 5. Speech-pressure waveforms (the upper of each pair) and associated speech processor outputs (lower of each pair) during the consonantal gesture of six intervocalic consonants via direct presentation. These are the same utterances used to construct Figure 4. To make the waveforms of roughly equal size here, [ata] was amplified by 6 dB, and [asa], [ava], [aza] and [aba] by 12 dB.

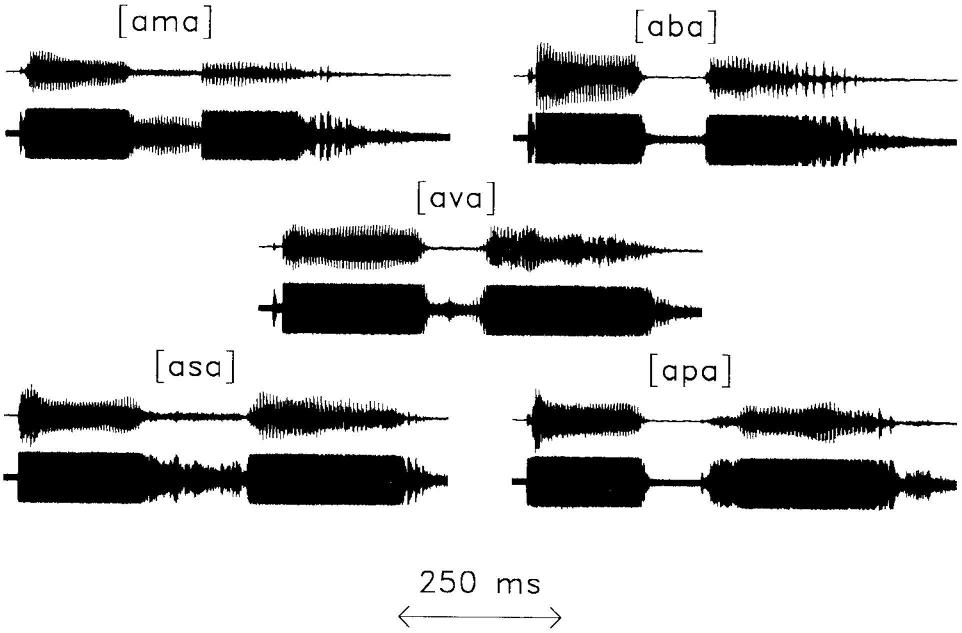

FIGURE 6. Examples of the speech-pressure waveforms and associated speech processor outputs for five intervocalic consonants through S2's processor via free-field presentation. These are the same tokens used in the construction of Figure 6, with which they can be compared.

of periodicity in the speech sound. Figure 5 shows, on expanded time scales, representative samples of the various possibilities, extracted from Figure 4. Both the nasal murmur of [m] and the sound associated with the closure for [b] showed the periodicity expected from vocal fold excitation. Similarly, the processor showed a clearly periodic output. Aperiodicity due to turbulence noise can be seen in the waveforms associated with the voiceless fricative [s], and the aspiration noise following the plosive release of [t].

Especially interesting are [v] and [z]. During the consonantal interval for these sounds, the excitation of the vocal tract is *mixed*, a combination of a periodic (laryngeal) source, and an aperiodic (turbulent) one. The relevant oscillograms of Figure 5 show that the balance of the mix is different for the two sounds, with aperiodicity dominating in [z] and the laryngeal periodicity dominating for [v] (reflecting the fact that the frication in [v] is at a considerably lower amplitude than that for [z]). This difference was reflected in the processor output, and so it might be expected that /z/ would have been labelled as a voiceless fricative more often than /v/, and that /v/ would have been frequently labelled as a voiced plosive. The confusion matrices (Figure 3) do not bear out this prediction. Although /v/, with its stronger voicing component was labelled a nasal slightly more frequently than /z/, it was also labelled as a voiceless fricative more often than /z/. This may be accounted for by the way the subjects' thresholds vary with frequency (Brimacombe, Edgerton, Doyle, Errat, & Danhauer, 1984), absolute sensitivity being better at lower frequencies. The lower frequency frication of /v/ might therefore have been perceptually more important than the higher amplitude (but higher frequency) frication of /z/.

Inspection of the processor output waveforms for a free-field connection showed important differences from the waveforms generated by a direct connection, most of which seem attributable to the free-field reproduction system (amplifier and loudspeaker). An example of one token from each of the five manner/voicing classes can be found in Figure 6. These may be compared to the corresponding waveforms in Figure 4, generated from identical tokens.

Note that for the voiced sounds [m] and [b], the amplitude of periodic excitation during consonantal closure was greatly reduced (or eliminated for [b]) in the free-field, both at the processor output and the acoustic pressure waveform generated near the microphone input. Because the sounds in these intervals were composed primarily of only the lowest few harmonics, this suggests that the loudspeaker/amplifier/room system had a high-pass characteristic that was attenuating low frequency components. Support for the view that high-frequency energy received relatively greater emphasis is seen also in the voiceless sounds [p] and [s], where for the processor outputs, the voiceless aperiodic excitation (at higher frequencies) was relatively greater in free-field than in direct recordings (especially for the [s]).[8] The output for

[8] Unlike the situation with the voiced sounds, here the difference between direct and free-field conditions is not very clear in the speech pressure waveforms. This may arise from the distortion of the amplitude envelopes of the vowels making it difficult to judge the relative level of periodic and aperiodic energy. In any case, as we shall see shortly, estimates of the amplitude response of the loudspeaker/amplifier/room system do indeed indicate a high-pass characteristic.

[v] showed both of these alterations at work. During the consonantal closure, periodic voicing dominated for direct presentation, whereas aperiodic frication dominated in the free-field.

Unfortunately, it was not possible to measure the amplitude response of the free-field system directly. Instead, it was estimated by dividing the long-term average spectrum of a set of free-field sounds by the long-term average spectrum of the same directly recorded sounds. Although there were peaks and valleys of 10 dB or more throughout the derived amplitude response, the overall trend was of a high-pass filter: a fairly flat region between 2 and 4 kHz; a shallow-sloping region (about 6–9 dB/octave) from 2 kHz down to about 200 Hz, and a more steeply-sloping region (about 27 dB/octave) from there down to 120 Hz.

Thus, the differences between speech processor output for free-field and direct connections seem most likely to have been due to the acoustic properties of the loudspeaker and/or room, as amplifiers usually have very flat amplitude responses.[9] The possible importance of this frequency shaping can be better appreciated for the particular case of the utterance [b]. The fundamental frequency for our female speaker fell to 180 Hz during the closure of [b], and this component, when measured directly from the video recorder, was at least 10 dB greater than any other. Therefore, its attenuation by the reproduction system (by nearly 20 dB in comparison to energy near 1 kHz) meant its effective disappearance from the acoustic waveform at the microphone to the speech processor, and necessarily, from the processor output.

Can this derived frequency shaping account for the perceptual data obtained? We mentioned above that the listeners labelled both voiceless plosives and fricatives more frequently as fricatives under free-field conditions, and this is consistent with the high frequency emphasis that the signals were undergoing. Unfortunately, it is not consistent with the results obtained for voiced fricatives, which would be expected to become more, rather than less, fricative-like. Further investigations are thus needed to develop a detailed explanation of the perceptual differences observed between direct and free-field presentations, even though the acoustical differences between the two conditions are reasonably well accounted for.

## *Question/Statement*

Waveforms from the speech processor to sentences from the question/statement task were also examined. As seen above in Figure 5, the processor output preserved the periodicity of the input waveform, as can also be seen for four vocalic segments in Figure 7 (direct connection).[10] All of our subjects were able to hear variations in fundamental frequency as changes in pitch, and use them in a linguistically appropriate way. They must therefore have been sensitive to the gross periodicity of the stimulating waveform. However, they seemed not to be sensitive to the temporal fine-structure differences between vowels, or at least were not able to use these features for vowel identification (as indicated by results on the MAC vowel subtest used by Edgerton, Prietto, & Danhauer, 1983[11]).

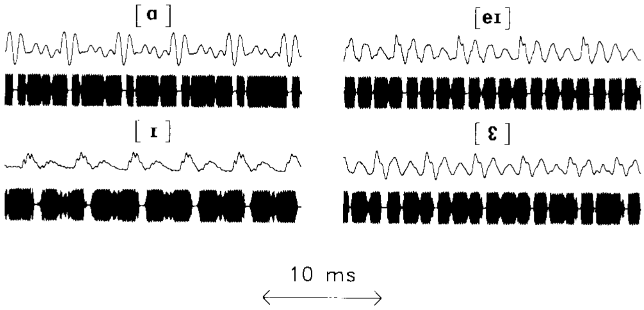

FIGURE 7. Examples of speech-pressure waveforms and speech processor outputs via direct presentation for four vocalic segments extracted from utterances in the question/statement task (female speaker). The lower row is excerpted from the sentence "It's down there?" with the left-most [ɪ] from the initial "It's" and the right-most [ɛ] from "there". The upper row is excerpted from the sentence "All the way" with the left-most [ɑ] from the initial "All" and the right-most [eɪ] the dipthong in "way".

The situation is more complicated for free-field presentation, as can be seen in Figure 8. The relative attenuation of the lower harmonics (and possible phase distortions) caused both the speech waveform as well as the processor output to be less clearly periodic than in the direct case. In the speech waveforms, for example, the

[9] It is, of course, possible that the microphone in the speech processor (which was bypassed in DT conditions) could be responsible for the filtering (as the same type of microphone was used for recording the acoustic pressure near the patient microphone). Fretz and Fravel (1985) claim, however, that "The microphone has a flat frequency response across the auditory range" (p. 16S).

[10] Wolf and Bilger (1977) have already shown that an early version of the House/3M device preserves the periodicity of an input vowel at its output. The output waveforms they display, however, are quite different from the ones we obtained in being much less clipped. It is not known whether this represents a change in the design of the speech processor, or simply results from different settings of the device.

[11] Doyle, Danhauer, and Edgerton (1986) attempted to determine the extent to which 15 users of the House/3M device were sensitive to differences among relatively steady-state portions of vowels excised from isolated words. They found levels of performance in an identification task that seem to be much superior to those of Edgerton, Prietto, and Danhauer (1983), a point not addressed in the later paper even though two authors are common to the two studies. Unfortunately, it is not clear from Doyle et al. whether the subjects are sensitive to spectral shape information in the form of the first formant frequency, or whether performance is completely dictated by differences in fundamental frequency. Further work, perhaps using vowels of identical fundamental, would help resolve this issue. Thus, there is only weak evidence that adult users of the House/3M device are able to use in a linguistic manner steady-state long-lasting spectral information contained in the fine time structure of the speech processor output. The results reported here for the identification of VCVs constitutes strong evidence that use of the House/3M device does not permit the linguistic use of dynamically-varying, or short duration, spectral features.

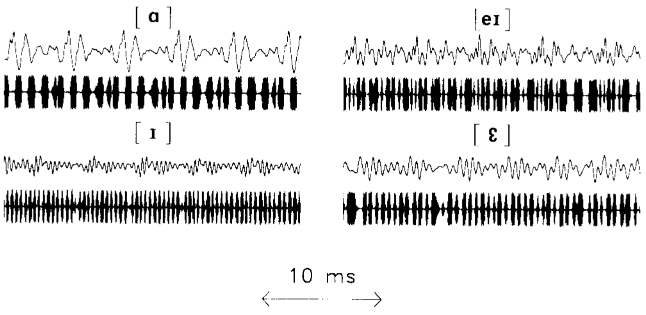

FIGURE 8. Examples of speech-pressure waveforms and speech processor outputs via free-field presentation for the same four vocalic segments used in Figure 7.

major peak of excitation that occurs for each vibratory cycle of the vocal folds was much more evident via a direct, rather than free-field, connection. Thus, although the free-field output waveforms were just as periodic as those from a direct connection, the periodicity was "disguised." Correspondingly, we might expect the percept of pitch to have been less salient in the free-field, a prediction not borne out by the small amount of testing done. Further aspects of this issue will be discussed below.

### *Outputs From Another Speech Processor*

All the waveforms we have discussed up until now were those obtained from S2's processor, representative of the behavior of 3 of the 4 processors. As noted above, S4 used her device on a setting that led to somewhat different output waveforms, as can be seen in Figure 9. Note first the less severe clipping of the vowels for S4's processor, especially for the two final vowels. Also, low amplitude events during consonantal closure (whether aperiodic for [k] or periodic for [n]) were better represented by S2's processor. Both these effects resulted from the fact that S4's processor sensitivity was set so as to work more often in the range below saturation (i.e., with the lowest sensitivity).

Not much is known about the extent to which subject performance varies with such differences in processor functioning. Edgerton and Brimacombe (1984) have presented evidence that more severe compression leads to improved performance in consonantal identification, yet S4's accuracy in all the tests we performed was more or less equivalent to that of the other 3 subjects. Of course, it may still be that a higher sensitivity setting (and more heavily clipped output waveforms) would improve her performance. If this were universally true, then it might be worthwhile to instigate programs of rehabilitation to ensure that patients use their devices on optimum settings.

It is also interesting to note that S4's pattern of errors in VCV identification did differ somewhat from that of the other 3 subjects. SINFA was applied to each of the subjects' summed confusion matrices for SP:DT[12] using the binary features of voicing, nasality, plosion, frication, and the four-valued feature, place. For S1, S2, and S3, SINFA identified four important features: nasality, voicing, frication, and place, in that order. For S4, the identified features in order of importance were plosion, voicing, nasality, and place. Inspection of the proportions of unconditional information transferred for the five manner/voicing classes helps to shed light on some of these findings. The relative unimportance for S4 of frication is reflected in the fact that she performed worst of the subjects in distinguishing voiceless fricatives from the other sounds (although she was the best in distinguishing voiced fricatives). On the other hand, her ability to distinguish voiced plosives was the best of the four, and voiceless plosives were only labelled correctly or as voiced plosives.

It is difficult to give a consistent account of these differences in performance based on the differences in the operation of the speech processors. After all, even subjects who have processors that behave in a similar manner can show differences in their performance (e.g., S2 and S3 were much worse in distinguishing voiced fricatives from the other sounds than S1). Above all, such detailed analyses should not blind us to the general

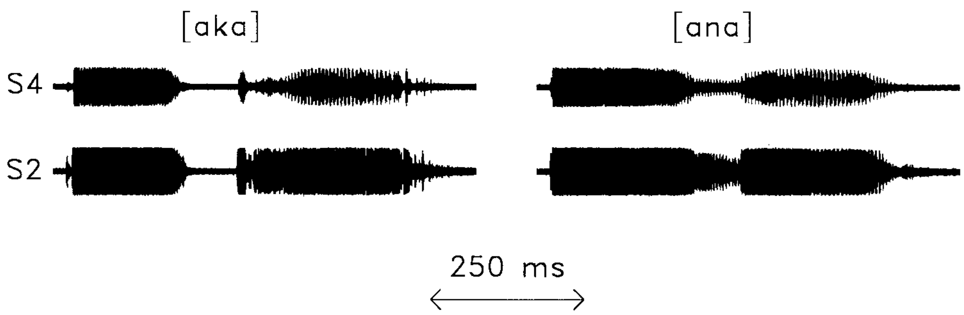

FIGURE 9. A comparison of the outputs of S4's and S2's processor (the latter used in Figures 4 to 8) for two intervocalic consonants presented in the free-field. The corresponding speech-pressure waveforms are *not* shown.

[12] Only the SP:DT matrices have been analyzed in this way because we wanted to focus attention on conditions where the role of auditory information was primary, and because relatively little data were collected in the SP:FF condition.

similarity of results across subjects. A feature based on the five manner/voicing categories accounts for 67%–86% (77% for S4) of the total information received by each subject in the condition SP:DT.

## CONCLUSIONS

Although there is a need to test a wider variety of subjects, it is clear that the House/3M device can provide important phonetic information, perhaps to a greater degree than has been appreciated before. This information takes two forms: segmental information as to the source of excitation (periodic vs. random) and its amplitude (especially presence vs. absence); and suprasegmental features of intonation. Because all of these are relatively invisible, they can complement the visual information available on the speaker's lips, and hence serve to significantly improve performance based on lipreading alone. Such improvements are not restricted only to closed-set tests like those used here. Robbins, Osberger, Miyamoto, Kienle, and Myres (1985) have shown that the House/3M device can significantly improve performance in a more natural task (connected discourse tracking), over that obtained by lipreading alone.

In the consonantal identification task, the ability to distinguish silence from high-amplitude periodicity from low-amplitude periodicity from aperiodic stimulation (and mixed periodic and aperiodic stimulation for some subjects) is sufficient to account for the main aspects of performance. Although much attention has been focused on amplitude envelope, per se, as a cue (e.g. Tyler, Tye-Murray, Preece, Gantz, & McCabe, 1987; Van Tasell, Soli, Kirby, & Widin, 1987), it is clear that this feature alone will not explain subject performance.

It is *not* that amplitude envelope features are unimportant—subjects clearly distinguished between silence and the presence of some sound, and between periodic stimulation of relatively high and low amplitudes (e.g., in distinguishing nasals from voiced plosives). However, at least as important to the subjects was whether the stimulation was periodic or aperiodic. If amplitude envelope were the crucial cue, then an [ɑsɑ] that had its frication artificially amplified to match the level of the surrounding vowel should be labelled as /ɑmɑ/ or /ɑnɑ/, a highly unlikely outcome. Supporting evidence for this assertion comes from the study of Edgerton and Brimacombe (1984). They asked users of the House/3M device to identify, by auditory means alone, a different set of VCVs (including /l/, /r/, /m/, /n/, /b/, /g/, /p/, /k/, /v/, /f/, /s/, and /ʃ/. Inspection of the output of the speech processor to /ɑʃɑ/, under conditions in which the outputs were highly clipped (as they typically were in our study), showed the medial fricative to have the same amplitude as the adjacent vocalic segments. Yet at least one subject (information about the performance of the other subjects is not sufficiently detailed) never misidentified this stimulus (see their Figure 6).

The popularity of the idea of amplitude envelope as a governing perceptual factor seems to rest, at least in part, on the idea that many patients receive *only* "time-intensity" information (another name for amplitude envelope) from an implant (see, for example, the discussions throughout Mecklenburg, 1986). In our experience (Fourcin et al., 1979), patients who are reasonably sensitive to time-intensity information (e.g., in temporal gap detection) readily distinguish between periodic and aperiodic stimulation (and also show some sensitivity to changes in fundamental frequency). From a theoretical viewpoint, too, all these abilities are dependent on temporal processing, so we would be surprised if they were not strongly correlated.

Taking the results of the prosodic and segmental tests together, it therefore appears that the primary features of the stimulation used by the subjects have to do with gross temporal fluctuations in amplitude envelope, and the degree of periodicity or aperiodicity in the signal. Although the stimulating waveform contains information about the spectral content of sounds (reflecting aspects of the temporal detail in the waveform), this spectral information is not able to be used in a linguistic way. That such temporal fine structure does not permit the identification of vowel quality or place of articulation in consonants is consistent with the finding that House/3M users cannot understand speech by auditory cues alone.

This is worthy of some explanation as it is well known that spectral shape distinctions *can* be conveyed through the temporal microstructure of the waveform applied to a single channel (e.g., Rosen & Ball, 1986; White, 1983), and discrimination of such contrasts has even been demonstrated in users of an early version of the House/3M device (Bilger, 1977). We, too, obtained evidence that temporal fine-structure in the stimulating waveform can be important in determining the nature of auditory percepts experienced by the subjects.

Recall that in the question/statement task, both triangle waves and 1-ms bursts of a 16-kHz carrier, when triggered at the appropriate rate, were able to successfully convey the question/statement distinction to the subjects. In other words, they conveyed melodic-pitch information. Yet subjects said that the two waveforms sounded quite different to them, even though they had the same periodicity. S3 reported that the triangles were at "a low pitch" whereas the 1-ms bursts of 16-kHz carrier were at "a medium or high-frequency pitch." It seems to us more likely that it was the timbre of the sounds that differed, as even normal listeners often use the terms "frequency" or "pitch" to describe differences in timbre rather than in fundamental frequency.

The inability of listeners to use this temporal fine-structure in a linguistic way may arise from at least three causes. First, it may be that only relatively large differences in fine-structure are discriminable, larger than are typically found in speech.

Secondly, there may be too much variability in the temporal structure of sounds effected by the acoustics of the surrounding environment (e.g., due to phase and amplitude distortions) to allow listeners to categorize sounds sufficiently consistently. Figures 7 and 8 show the large differences in stimulating waveform for the same

TABLE 5. Collapsed confusion matrices obtained from 1 subject wearing the Nucleus device, and one wearing the House/3M device, both in conditions SP:FF and listening to the same stimulus tapes.

| | | *Condition SP:FF* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *Nucleus* | | | | | *House/3M - S1* | | | | | | |
| | | *Response* | | | | | *Response* | | | | | | |
| | | *mn* | *vz* | *bdg* | *ptk* | *fs* | *mn* | *vz* | *bdg* | *ptk* | *fs* | | |
| Stimuli | mn | 100 | | | | | 56 | 25 | 13 | | 6 | mn | Stimuli |
| | vz | 50 | | 25 | 13 | 13 | 13 | 56 | 25 | | 6 | vz | |
| | bdg | | 8 | 92 | | | 8 | 54 | 29 | 4 | 4 | bdg | |
| | ptk | | 8 | 50 | 33 | 8 | | | 8 | 83 | 8 | ptk | |
| | fs | | 50 | 13 | 13 | 25 | | | | 6 | 94 | fs | |
| | | 19/48 correct = 40% | | | | | 34/96 correct = 35% | | | | | | |
| | | % information | | | | | % information | | | | | | |
| | | Voicing = 12% Place = 45% Manner = 45% | | | | | Voicing = 67% Place = 12% Manner = 22% | | | | | | |

*Note.* Table format as for Table 3.

vowel via direct and free-field presentations. Note that the gross temporal features being used by the subjects (periodicity vs. aperiodicity; high amplitude sound vs. low amplitude sound vs. silence) are less variable in this way.

Finally, it may be that the speech processor alters the natural temporal structure of speech sounds so as to lead to a stimulating waveform that elicits percepts so different from those experienced in normal hearing that the new categories cannot be learned. Reports that a small number of children (4% of the total reported) using the House/3M device can understand speech without lipreading (Luxford, Berliner, Eisenberg, & House, 1987) may imply that the last factor is crucial, as children presumably retain a much greater capacity for learning. However, the fact that so few children appear to develop these extraordinary skills supports the idea that the spectral shape information in the stimulating waveform is extremely difficult to extract.

For the adult user, there is no doubt that it would be advantageous to explicitly control the temporal fine-structure of the stimulation to signal some broad classes of spectral shape, rather than letting the stimuli reflect aspects of spectral structure in a form that seems unusable at present.

Many other questions remain unanswered. For instance, in the perception of voiceless plosives, it is not clear which of the likely distinguishing cues (silent interval, plosive burst, or aspiration) are primary, or whether all can be used. Questions like these may be answered by experiments with manipulated natural stimuli in which, for example, plosive bursts or intervals of aspiration are deleted, or emphasized.

A crucial issue is the extent to which the perception of voice pitch contours depends on the detailed shape of the stimulating waveform (e.g., triangular pulses vs. vowels). This is especially important for two reasons. First, it appears that suprasegmental intonation is a more important aid to lipreading connected discourse than the segmental cues that indicate presence or absence of voicing (Rosen et al., 1980). Second, there is some evidence that changes in signal processing (in terms of simplifying the input waveform) could improve the perception of voice pitch changes.

Through such studies, it may be possible to match the stimulation to the patient more accurately than it is now, ensuring for example, that voiced fricatives and voiced plosives are more effectively distinguished. Such explicit matching between the subjects' needs and abilities are more readily effected with a speech pattern (feature extracting) approach than the purely pragmatic one that seems to have determined the current design of the House/3M device. A speech pattern processor that included fricative information, better signalling of voice fundamental frequency, and was more resistant to disruption by background noise than the current House/3M/patient system,[13] could be of greater benefit than the current speech processor.

In any case, it appears that, as it stands, the House/3M device can provide important cues to intonation, manner, and voicing that are significant aids to lipreading. That such important information can be readily signalled by temporal features in a single channel suggests that it is the minimum information that any implant system should convey.

Interestingly, Rosen (1987) has found that multichannel implant systems do not necessarily signal this basic information very well. Table 5 compares collapsed con-

[13] Background noise is, of course, a perennial problem for users of nearly all auditory prostheses. Unfortunately, almost nothing is known about the levels of noise that still permit efficient use of the House/3M device. Although a speech pattern approach, with explicit extraction of important features, theoretically allows the provision of noise-free signals to the listener, current feature extraction circuits are themselves easily disrupted by background noise. Future improvements in speech analysis capabilities may radically alter this situation.

fusion matrices for S1 (the best performer in the present study), and for one subject wearing the multi-channel Nucleus device (the best performer of 5 subjects tested[14]). Both listened to the same stimulus tapes in condition SP: FF. Detailed comparisons are unwise because, although the stimulus materials were identical, different reproduction systems were used for the two tests. In particular, the relatively poor perception of manner by S1 probably results from the distortions introduced by the free-field system (see Table 4 and the previous section). Even so, note that the Nucleus device led to much more place information being transmitted whereas the House/3M system more successfully conveyed voicing information.[15]

Although systems that are better at transmitting place information at the cost of basic voicing/manner features may lead to some small benefits in understanding speech without lipreading compared to systems that transmit voicing/manner cues best, they may prove less useful for many implant users trying to understand speech in realistic situations (i.e., when lipreading is necessary). Designers of complex multi-channel systems still have important lessons to learn "in time" from relatively simple but effective single-channel devices.

## ACKNOWLEDGMENTS

Grateful appreciation is extended to: the audio-visual department at the House Ear Institute who handled all the technical aspects of the recordings; at University College London: David Smith and Ian Howard who helped in the analysis of fundamental frequency contours; Mark Huckvale, who provided much assistance in constructing the figures; David Smith for programming a complete package for analyzing the results of the intervocalic consonant test and Ginny Ball for using it so assiduously. Moshe Yuchtman kindly provided the programs for SINFA at short notice. Welcome comments on the manuscript were received from Ginny Ball, Andrew Faulkner, Adrian Fourcin, Mary Joe Osberger, and David Smith. Most of all, we are thankful to the four subjects who gave freely of their time and effort.

SR and JRW were supported in this work by the Medical Research Council of the United Kingdom. JB and BE were with the House Ear Institute.

[14] The results of this user of the Nucleus device are fairly typical of the best performances evidenced by Nucleus users using a speech processor that codes fundamental frequency and the frequency and amplitude of the second formant. Dowell, Tong, Blamey, and Clark (1985), for instance, report the results of 4 subjects identifying the same 12 intervocalic consonants (albeit from a different speaker). The highest score was 56% overall correct with 13% of voicing, 45% of place, and 45% of manner information transmitted correctly.

[15] Up until now, we have only discussed performance by users of a Nucleus speech processor that incorporates information about fundamental frequency and the second formant. There is evidence that a newer speech coding scheme for the Nucleus device, which also signals information about the frequency and amplitude of the first formant (Dowell, Seligman, Blamey, & Clark, 1987), leads to better transmission of voicing information (Blamey, Dowell, Brown, Clark, & Seligman, 1987). Presumably, this improved performance results from subjects being sensitive to an overall spectral balance in the signal, weighted to the low frequencies for voiced sounds, and to the high frequencies for voiceless ones. It may still be worthwhile to ensure that segmental voicing information is coded reliably in the temporal pattern of stimulation.

## REFERENCES

ABBERTON, E., & FOURCIN, A. J. (1984). Electrolaryngography. In C. Code & M. Ball (Eds.), *Experimental clinical phonetics* (pp. 62–78). London: Croom Helm.

ABBERTON, E., FOURCIN, A. J., ROSEN, S., WALLIKER, J. R., HOWARD, D. M., MOORE, B. C. J., DOUEK, E. E., & FRAMPTON, S. (1985). Speech perceptual and productive rehabilitation in electrocochlear stimulation. In R. A. Schindler & M. M. Merzenich (Eds.), *Cochlear implants* (pp. 527–537). New York: Raven Press.

BILGER, R. C. (1977). Evaluation of subjects presently fitted with implanted auditory prostheses. *Annals of Otology, Rhinology and Laryngology, 86*(Suppl. 38).

BLAMEY, P. J., DOWELL, R. C., BROWN, A. M., CLARK, G. M., & SELIGMAN, P. M. (1987). Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *Journal of the Acoustical Society of America, 82*, 48–57.

BRIMACOMBE, J. A., EDGERTON, B. J., DOYLE, K. J., ERRAT, J. D., & DANHAUER, J. L. (1984). Auditory capabilities of patients implanted with the House single-channel cochlear implant. *Acta Oto-laryngologica* (Stockholm), (Suppl. 411), 204–216.

DANLEY, M. J., & FRETZ, R. J. (1982). Design and functioning of the single-electrode cochlear implant. *Annals of Otology, Rhinology and Laryngology, 91*(Suppl. 91), 21–26.

DOWELL, R. C., MARTIN, L. F. A., TONG, Y. C., CLARK, G. M., SELIGMAN, P. M., & PATRICK, J. F. (1982). A 12-consonant confusion study on a multiple-channel cochlear implant patient. *Journal of Speech and Hearing Research, 25*, 509–516.

DOWELL, R. C., SELIGMAN, P. M., BLAMEY, P. J., & CLARK, G. M. (1987). Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. *Annals of Otology, Rhinology & Laryngology, 96*(Suppl. 128), 132–134.

DOWELL, R. C., TONG, Y. C., BLAMEY, P. J., & CLARK, G. M. (1985). Psychophysics of multiple-channel stimulation. In R. A. Schindler & M. M. Merzenich (Eds.), *Cochlear implants* (pp. 283–290). New York: Raven Press.

DOYLE, K. J., DANHAUER, J. L., & EDGERTON, B. J. (1986). Vowel perception: Experiments with a single-electrode cochlear implant. *Journal of Speech and Hearing Research, 29*, 179–192.

EDGERTON, B. J., & BRIMACOMBE, J. A. (1984). Effects of signal processing by the House-3M cochlear implant on consonant perception. *Acta Oto-laryngologica* (Stockholm), (Suppl. 411), 115–123.

EDGERTON, B. J., DOYLE, K. J., BRIMACOMBE, J. A., DANLEY, M. J., & FRETZ, R. J. (1983). The effects of signal processing by the House-Urban single-channel stimulator on auditory perception abilities of patients with cochlear implants. *Annals of the New York Academy of Sciences, 405*, 311–322.

EDGERTON, B. J., PRIETTO, A., & DANHAUER, J. L. (1983). Cochlear implant patient performance on the MAC battery. *Otolaryngologic Clinics of North America, 16*, 267–280.

FOURCIN, A. J., ROSEN, S. M., MOORE, B. C. J., DOUEK, E. E., CLARKE, G. P., DODSON, H., & BANNISTER, L. H. (1979). External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings. *British Journal of Audiology, 13*, 85–107.

FRETZ, R. J., & FRAVEL, R. P. (1985). Design and function: A physical and electrical description of the 3M House cochlear implant system. *Ear and Hearing, 6*(No. 3 Suppl.), 14S–19S.

GRANT, K. W., ARDELL, L. H., KUHL, P. K., & SPARKS, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77,* 671–677.

HOCHMAIR, E. S., & HOCHMAIR-DESOYER, I. J. (1985). Aspects of sound signal processing using the Vienna intra- and extra-cochlear implants. In R. A. Schindler & M. M. Merzenich (Eds.), *Cochlear implants* (pp. 101–110). New York: Raven Press.

HOWARD, D. M., & FOURCIN, A. J. (1983). Instantaneous voice period measurement for cochlear stimulation. *Electronics Letters, 19,* 776–778.

LEDER, S. B., SPITZER, J. B., MILNER, P., FLEVARIS-PHILLIPS, C., RICHARDSON, R., & KIRCHNER, J. C. (1986). Reacquisition of contrastive stress in an adventitiously deaf speaker using a single-channel cochlear implant. *Journal of the Acoustical Society of America, 79,* 1967–1974.

LISKER, L., & ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops. *Word, 20,* 384–422.

LUXFORD, W. M., BERLINER, K. I., EISENBERG, L. S., & HOUSE, W. F. (1987). Cochlear implants in children. *Annals of Otology, Rhinology and Laryngology,* 96(Suppl. 128), 136–138.

MECKLENBURG, D. J. (Ed.). (1986). Cochlear implants in children. *Seminars in Hearing, 70.*

MILLER, G. A., & NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352. (Reprinted in I. Lehiste (Ed.), (1967). *Readings in acoustic phonetics.* Cambridge, MA: MIT Press.)

OWENS, E., KESSLER, D. K., TELLEEN, C. C., & SCHUBERT, E. D. (1981). *The minimal auditory capabilities battery* (Instruction manual). St. Louis: Auditec.

RISBERG, A. (1974). The importance of prosodic speech elements for the lipreader. *Scandinavian Audiology,* (Suppl. 4), 153–164.

ROBBINS, A. M., OSBERGER, M. J., MIYAMOTO, R. T., KIENLE, M. L., & MYRES, W. A. (1985). Speech-tracking performance in single-channel cochlear implant subjects. *Journal of Speech and Hearing Research, 28,* 565–578.

ROSEN, S. (1987). The perception of consonants with single- and multi-channel cochlear implants. *Acoustics Bulletin, 12,* 8 (Edinburgh, United Kingdom: Institute of Acoustics).

ROSEN, S., & BALL, V. (1986). Speech perception with the Vienna extra-cochlear single-channel implant. *British Journal of Audiology, 20,* 61–63.

ROSEN, S., FOURCIN, A. J., ABBERTON, E., WALLIKER, J. R., HOWARD, D. M., MOORE, B. C. J., DOUEK, E. E., & FRAMPTON, S. (1985). Assessing Assessment. In R. A. Schindler & M. M. Merzenich (Eds.), *Cochlear implants* (pp. 479–498). New York: Raven Press.

ROSEN, S., FOURCIN, A., ABBERTON, E., WALLIKER, J., DOUEK, E., MOORE, B., FRAMPTON, S., & HOWARD, D. (1983). Assessment of speech receptive and productive ability with electrically stimulated hearing. *11e Congrés International d'Acoustique: Vol. 4 (11th International Congress on Acoustics), Revue d'acoustique* (hors série), 297–300.

ROSEN, S., FOURCIN, A. J., & MOORE, B. C. J. (1980). Lipreading connected discourse with fundamental frequency information. *British Society of Audiology Newsletter,* (Summer Issue, August), 42–43.

ROSEN, S., FOURCIN, A. J., & MOORE, B. C. J. (1981). Voice pitch as an aid to lipreading. *Nature, 291,* 150–152.

ROSEN, S., MOORE, B. C. J., & FOURCIN, A. J. (1979). Lipreading with fundamental frequency information. *Proceedings of the Institute of Acoustics Autumn Conference 1979,* (Paper 1A2, pp. 5–8). Edinburgh, UK: Institute of Acoustics.

TYLER, R. S., GANTZ, B. J., MCCABE, B. F., LOWDER, M. W., OTTO, S. R., & PREECE, J. P. (1985). Audiological results with two single channel cochlear implants. *Annals of Otology, Rhinology and Laryngology, 94,* 133–139.

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. *Annals of Otology, Rhinology and Laryngology,* 96(Suppl. 128), 141–144.

VAN TASSELL, D. J., SOLI, S. D., KIRBY, V. M., & WIDIN, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America, 82,* 1152–1161.

WANG, M. D., & BILGER, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America, 54,* 1248–1266.

WHITE, M. W. (1983). Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes. *Annals of the New York Academy of Sciences, 405,* 348–359.

WOLF, R. V., & BILGER, R. C. (1977). Electroacoustic measures of present prostheses. In R. C. Bilger (Ed.), Evaluation of subjects presently fitted with implanted auditory prostheses. *Annals of Otology, Rhinology and Laryngology, 86*(Suppl. 38).

Received July 9, 1987
Accepted May 17, 1988

Requests for reprints should be sent to Stuart Rosen, University College London, Department of Phonetics and Linguistics, 4 Stephenson Way, London NW1 2HE England.