Auditory/phonetic categorization with the Symbion multichannel cochlear implant

Michael F. Derman
Arizona State University, Tempe, Arizona 85287

Maureen T. Hannley
Stanford University Medical Center, Palo Alto, California 94305

Geary A. McCandless
University of Utah, Salt Lake City, Utah 84103

Luke M. Smith
Symbion, Inc., 825 North 300 West, Salt Lake City, Utah 84103

(Received 27 April 1987; accepted for publication 4 April 1988)

The phonetic identification ability of an individual (SS) who exhibits the best, or equal to the best, speech understanding of patients using the Symbion four-channel cochlear implant is described. It has been found that SS: (1) can use aspects of signal duration to form categories that are isomorphic with the phonetic categories established by listeners with normal auditory function; (2) can combine temporal and spectral cues in a normal fashion to form categories; (3) can use aspects of fricative noises to form categories that correspond to normal phonetic categories; (4) uses information from both F1 and higher formants in vowel identification; and (5) appears to identify stop consonant place of articulation on the basis of information provided by the center frequency of the burst and by the abruptness of frequency change following signal onset. SS has difficulty identifying stop consonants from the information provided by formant transitions and cannot differentially identify signals that have identical F1's and relatively low-frequency F2's. SS's performance suggests that simple speech processing strategies (filtering of the signal into four bands) and monopolar electrode design are viable options in the design of cochlear prostheses.

PACS numbers: 43.71.Ky, 43.71.Cq, 43.63.Mb

INTRODUCTION

In this article, we describe the phonetic identification ability of a patient using a Symbion multichannel cochlear implant (Eddington, 1980). The patient's speech understanding is the best, or equal to the best, of the approximately 100 listeners using the Symbion device. Thus his performance may represent asymptotic use of information provided by an implant of this design.

We have used three strategies to assess the acoustic information used by our patient for phonetic identification. One strategy was to synthesize signals in such a manner that the signals formed a continuum along a single, or small number, of acoustic dimensions (see also Merzenich, 1985; Hood et al., 1987). We assumed that, if our patient could differentially identify signals from the continuum, then we would have evidence that the implant transmits information about the dimension which underlies the continuum in a form that the listener could use in identification.

We constructed a set of continua that sampled manner, place and voicing distinctions, and the types of acoustic properties that convey phonetic distinctions. Four continua were based on the durations of acoustic events. Four others were based on the frequency composition of acoustic events.

The continua based on duration of acoustic events included:

(a) A "chop" to "shop" continuum in which friction duration varied over the range 60–140 ms.
(b) A "slit" to "split" continuum in which the silent interval between the "s" and "lit" varied over the range 20–120 ms.
(c) Two "ga" to "ka" continua in which the interval between burst and onset of voicing (the voice onset time or VOT) varied over the range 0–45 ms. The stimuli in the two continua differed by approximately 120 Hz in the starting frequency of the first formant (F1). We included these continua in our test battery in order to assess the interaction of temporal and spectral cues in determining voicing. For normal listeners, a stimulus with low-F1 starting frequency will elicit more voiced responses than a stimulus with higher F1 starting frequency, presumably because a low F1 signals a more closed vocal tract (Lisker, 1975).

The continua based on differences in frequency composition of acoustic events included:

(d) An /s/-/z/ continuum in which the first formant (F1) and the second formant (F2) covaried over the range 420–580 Hz and 1780–1620 Hz, respectively.
(e) An /s/-/a/ continuum in which the first formant was fixed at 680 Hz and F2 varied over the range 880–1180 Hz.
(f) A "save" to "shave" continuum in which the center
frequency of the fricative noise varied over the range approximately 4500–3000 Hz.

(a) A "ba" to "da" continuum in which the starting frequency of F2 and F3 varied over the ranges 900–1700 Hz and 2000–2800 Hz, respectively. The stimuli in this series were created without release bursts. Thus the phonetic distinction was carried solely by the formant transitions.

Our goals in using these continua were to determine: (1) whether the listener could identify signals at the end points of the continua; (2) whether identification performance varied systematically with small changes in the acoustic input; and (3) whether the categories defined by the identification functions conformed to the phonetic categories established by listeners with normal auditory function.

A second strategy to assess the acoustic information used in phonetic identification is to remove a putative acoustic cue from a signal and then to compare the identification of the altered signal with that of the unaltered signal. We adopted this strategy to assess the use of F1 and higher formants in the identification of 12 vowels in "bVt" format. Identification accuracy was assessed for unaltered signals, for signals with only F1 present, and for signals with only F2 and higher formants present.

A third strategy to assess the acoustic cues used by a listener in phonetic identification is to create signals with conflicting cues to identity. For example, stop consonants can be created with formant transitions appropriate for one place of articulation and with spectral shape at onset (or spectral change from burst to voicing onset) appropriate for another place (see, for example, Blumstein et al., 1982; Lahiri et al., 1984; Walley and Carrell, 1983). A listener's identification of stimuli of this kind provides evidence about the relative salience of the cues to phonetic identity. For the present experiment, we created stimuli with three potential cues to stop consonant identity: spectral shape at signal onset (Blumstein and Stevens, 1979); abruptness of frequency change following onset (Kewley-Port, 1983); and onset frequency and direction of change of formant transitions (Lieberman et al., 1954). Our interest was the relative salience of the several cues to stop consonant identity for our patient.

I. METHODS
A. Subjects

Ten undergraduate volunteers at Arizona State University served as control subjects. All offered a verbal report of normal hearing.

The experimental subject (SS) was a 35-year-old male whose hearing loss was first diagnosed following a case of mumps at age 6 years. The patient reports that, at age 6 years, thresholds in the left ear averaged approximately 50 dB HL, while those in the right ear averaged 75 dB. From age 6–21 years, the thresholds in both ears worsened, reaching 90 dB in the left ear and 100 dB in the right ear. From age 21–25 years, thresholds in both ears fluctuated widely. At age 25, thresholds in the left ear were unmeasurable. During the period from age 25–33 years, the patient reports that he relied on lipreading exclusively for purposes of speech understanding and that a hearing aid did not improve speech understanding.

Immediately prior to surgery in April 1984, thresholds (in HL) for the right ear were 90 dB at 0.25 kHz, 100 dB at 0.5 kHz, 110 dB at 1 kHz, 100 dB at 2 kHz, 105 dB at 3 kHz, and no response at 4 and 8 kHz. Speech discrimination at 110 dB was 0 percent correct. In the left ear, signals were undetectable at 110 dB. The left ear was implanted.

Twelve months after implantation, SS achieved the following scores on the minimal auditory capabilities (MAC) battery (Owens et al., 1981): initial consonants = 94 percent correct; vowels = 97 percent correct; final consonants = 96 percent correct; monosyllabic words = 62 percent correct. SS's monosyllable recognition scores have ranged from 44–73 percent correct over test sessions.

On a recent test of consonant identification in "aCa" format (16 items), SS achieved a score of 81 percent correct. On a test of vowel identification in "bVt" format (12 items), SS achieved a score of 79 percent correct (see Sec. II). Both tests were administered as auditory tests, i.e., without lip-reading.

SS had participated as a listener in many psychoacoustic and speech perception experiments prior to participation in the present research. He is identified as U2 in Gantz et al. (1987).

B. Implant design

The Symbion implant consists of: (1) six monopolar electrodes implanted in the scala tympani with remote reference; (2) a percutaneous pedestal to which the electrode wires are attached; and (3) a portable speech processing and electrode stimulation system (Eddington, 1980, 1983). The most apical electrode is located about 22 mm from the round window. The electrodes are spaced at 4-mm intervals. Usually, the four most apical electrodes are activated. Each electrode is driven by a signal derived from the input signal after bandpass filtering. The center frequencies of the filters for channels 1–4 (most apical to most basal electrodes) are 0.5, 1, 2, and 4 kHz.

C. Stimuli

1. Temporal continua

(a) "shop"—"chop." To create a continuum in which the duration of a steady-state signal varied, the word shop, spoken by a male, was first sampled at 20 kHz and then stored in digital form in computer memory. Using a waveform editor, the "sh" noise was removed from the signal and stored independently of the vocalic portion. The rise time of the noise was edited to 50 ms by trimming the onset. The fall time was edited to 10 ms. The peak energy in the fricative noise was at 2240 Hz, as determined by LPC analysis. "Center cuts" were then made in the noise so that when the initial and final portions of the noise were rejoined, nine stimuli were created with durations of 60–140 ms in 10-ms intervals. By removing portions from the middle of the noise segment, we left its onset and offset unchanged. The nine fricative noises were then rejoined with the vocalic portion of the word to form a set of stimuli in which only the noise portion varied. The stimuli with brief fricative noise were intended to be heard as chop, while those with longer noise were intended to be heard as shop.
(b) “slit”–“split.” To create a continuum in which the duration of a silent interval varied, the word split, spoken by a male, was first sampled at 20 kHz and stored in digital form. The 250-ms “s” noise was then removed from the vocalic portion of the utterance. The peak energy in the noise was at 3.8 kHz and was 4 dB less than peak vowel energy. The vocalic portion of the utterance was edited to remove the low-amplitude burst and first voiced pulse. Intervals of silence of 20, 40, 50, 60, 70, 80, 90, 100, and 120 ms were then inserted between the noise and vocalic portion of the utterance. The stimuli with brief silent intervals were intended to be heard as slit, while those with longer silent intervals were intended to be heard as split.

(c) “ga”–“ka.” Two ga-ka continua were constructed. To generate the stimuli, ga was first synthesized with five formants using the KLATT algorithm (these stimuli were modeled after those in Repp, 1983, Experiment 5). The synthesizer was configured in parallel mode. The steady-state values of $F_1$–$F_5$ were 765, 1230, 2527, 3300, and 3850 Hz. The starting frequencies of $F_1$, $F_2$, and $F_3$ were 407, 1764, and 2200 Hz. Transition duration was 50 ms. All transitions were created by linear interpolation. The frequencies of $F_4$ and $F_5$ did not vary over time. A 10-ms burst centered at 1763 Hz preceded the voiced formants. The fundamental frequency was 120 Hz for the first 155 ms and then decreased linearly over the next 150 ms to a value of 98 Hz. Stimulus duration was 305 ms. To create the nine test stimuli, the time between the offset of the burst and the onset of voicing was varied over the range 0–45 ms in 5-ms steps. As the VOT increased, so also did the starting frequency of $F_1$ (range 407–693 Hz). For the duration of the delay in voicing onset, the upper resonators were excited by hiss to simulate aspiration. Post synthesis measurements indicated that the rms level of the aspiration was approximately 20 dB relative to the voiced portion.

To create the stimuli for the second ga–ka continuum, the 0-ms VOT stimulus was first altered. $F_1$ began at 279 Hz (128 Hz lower than for the 0-ms VOT stimulus of the first series). The duration of the $F_1$ transition was lengthened to 70 ms. The steady-state frequency of $F_1$ was the same as for the stimuli of the first continuum. The remaining eight stimuli were created in the manner of the first series. For each VOT of the second series, the $F_1$ onset frequency was approximately 120 Hz lower than the corresponding $F_1$ onset frequency of the first series.

2. Spectral continua

(a) /s/-/ʃ/. To create a continuum in which $F_1$ and $F_2$ covaried, the vowel /s/ was first synthesized with five formants by the KLATT algorithm. The synthesizer was configured in the cascade mode. The frequency of $F_1$ was 420 Hz and that of $F_2$ was 1780 Hz. $F_3$, $F_4$, and $F_5$ were at 2570, 3300, and 3850 Hz, respectively. Stimulus duration was 150 ms with 25-ms rise time and 50-ms fall time. The fundamental frequency began at 105 Hz, increased to 125 Hz over 25 ms, and then decreased to 80 Hz over 125 ms. All changes in $F_0$ were made by linear interpolation. A total of nine stimuli were created by varying $F_1$ and $F_2$, in 20-Hz steps, over the range 420–580 Hz and 1780–1620 Hz, respectively. Overall signal level varied less than 1 dB over the stimulus range.

(b) /ʃ/-/s/. To create a continuum in which $F_1$ was fixed and $F_2$ varied, /ʃ/ was synthesized with five formants in parallel mode using the KLATT algorithm. $F_1$ was at 620 Hz, $F_2$ was at 880 Hz, $F_3$ was at 2500 Hz, $F_4$ was at 3300 Hz and $F_5$ was at 3850 Hz. Stimulus duration was 180 ms with a 25-ms rise time and a 50-ms fall time. $F_0$ began at 105 Hz, increased to 125 Hz over 25 ms, and then decreased to 80 Hz over the final 125 ms. All changes were made by linear interpolation. A total of nine stimuli were made by varying $F_2$ frequency. The second stimulus in the series was at 900 Hz. The others were spaced in 40-Hz increments. Signal level varied less than 1.5 dB over the stimulus range.

(c) “save”–“shave.” To create a continuum in which the location of fricative energy varied, the word save, spoken by a male, was digitized at a 10-kHz sampling rate and stored in computer memory. The vocalic portion of the word was edited and stored. Eight fricative noises were then constructed using the KLATT algorithm. Each noise was generated by controlling two poles of the synthesizer. For the extreme “s”-like noise, the poles were at 4999 and 3917 Hz. For the extreme “sh”-like noise, the poles were at 3915 and 2197 Hz. Intermediate stimuli were constructed with poles between the two extremes. The length of the fricative noise was 210 ms. The signals were digitally edited so that the SPL of the noises varied by less than 1.5 dB over the stimulus range. Finally, the eight noises were recombined with the natural voice vocalic portion of “save” to form the eight test stimuli. Total stimulus duration was 924 ms.

(d) “ba”–“da.” To create this continuum, ba was first synthesized with five formants by the KLATT algorithm. The synthesizer was configured in cascade mode. The steady-state frequencies of $F_1$–$F_5$ were 720, 1240, 2500, 3600, and 4500 Hz, respectively. $F_1$ began at 520 Hz and increased to 720 Hz over 15 ms, $F_2$ began at 900 Hz and increased to 1240 Hz over 40 ms, and $F_3$ began at 2000 Hz and increased to 2500 Hz over 40 ms. All transitions were created by linear interpolation. Stimulus duration was 200 ms with a 30-ms rise time and a 50-ms fall time. Eight stimuli were created by varying the onset frequency of $F_2$ over the range 900–1700 Hz in approximately 110-Hz increments and by covarying the onset frequency of $F_3$ over the range 2000–2800 Hz in 100-Hz increments.

3. Vowels in “bVt” format

To assess the cues used in vowel identification, the words “beet, bit, bait, bet, bought, but, boot, Bert, bout, bite, boat” were synthesized with five formants using the KLATT algorithm. The synthesizer was configured in parallel mode with cascade adjustment. Each stimulus was composed of a 5-ms /b/ burst, a 5-ms silent interval, 30 ms /b/ transitions, 90-ms vocalic nucleus, 50-ms /t/ transitions, 80 ms of silence, and a final /t/ burst of 50 ms. The center frequencies of the first three formants are shown in Table I. $F_1$, $F_2$, and $F_3$ were varied in frequency for some of the vowels in order to improve their intelligibility. The F0 track was the same for all signals. The peak SPL, measured at the level of the subject’s input microphone, varied less than 1.5 dB across the 12 signals. The amplitude envelopes of the signals varied slightly.

Two additional sets of bVt stimuli were created by eliminating a formant or formants from the synthesis declarations. For one set, the stimuli were synthesized with labial release burst, first formant, and alveolar release burst (F1-only condition). In the other set, the stimuli were synthesized with labial release burst, F2–F5, and alveolar release burst (F1–F2). All three formants were present in stimuli consisting of the three formants (e.g., the stimulus with /b/ transition cues). To create the gill stimulus with abrupt frequency change following onset, the desired length of waveform (18 ms) was removed from the segment following the release burst; i.e., the burst was not removed. Stimuli in which both spectral shape and abruptness of frequency change cues conflicted with formant transition cues were created by keeping the spectral shape constant for the duration of the burst and aspiration segment.

D. Procedure

Each test was presented as a forced-choice identification test with the alternatives listed in orthographic form at the top of the printed test sheet. The test sequence for each continuum consisted of eight tokens of each stimulus presented in random order. The interstimulus interval was 3.5 s. Each test sequence was preceded by a familiarization sequence in which the extreme stimuli from the continuum were presented as stimulus pairs. Eight repetitions of the pairs were presented. The subject was then asked whether he would like to hear the pairs a second time. For SS only the ba–da and slit–split stimuli were presented a second time. The test sequence followed immediately. The subject was instructed to write the identity of each test stimulus on the answer sheet.

The test for the bVt stimuli consisted of a randomized sequence of ten tokens of each of the stimuli. The familiarization sequence consisted of three repetitions of the 12 vowels. For this sequence, stimulus identity was indicated on the answer sheet. The interstimulus interval was 3.5 s.

The test for the conflicting cue stimuli consisted of a randomized sequence of eight tokens of each of the stimuli. The familiarization sequence consisted of five repetitions of the bill, /d/ and gill stimuli with a normal complement of cues. For this sequence, stimulus identity was indicated on the answer sheet.

All listening tests were conducted in a sound attenuating booth. SS sat approximately 3 m from a loudspeaker. His implanted ear was oriented toward the loudspeaker. The signal presentation level was 75 dB SPL (C weighted). The normal-hearing listeners were tested using TDH-39 headphones at the same signal level.

II. RESULTS AND DISCUSSION

A. Continua

In Figs. 1–8, the identification responses (± 1 standard deviation) of the normal-hearing listeners are shown by the dotted functions.

1. Temporal continua

SS’s identification of the signals from the shop–chop series is shown in Fig. 1. Signals near the extremes of the continuum were labeled without variation. This suggests that the identity of the signals was unambiguous. The location of the category boundary (the point of 50% identification) was

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>beet</td>
<td>313–200</td>
<td>2048</td>
<td>2695</td>
</tr>
<tr>
<td>bit</td>
<td>390–460</td>
<td>1756–1556</td>
<td>2456–2536</td>
</tr>
<tr>
<td>bet</td>
<td>546</td>
<td>1610</td>
<td>2539</td>
</tr>
<tr>
<td>bat</td>
<td>703</td>
<td>1562</td>
<td>2500</td>
</tr>
<tr>
<td>bought</td>
<td>625–650</td>
<td>859–910</td>
<td>2700–2730</td>
</tr>
<tr>
<td>but</td>
<td>625</td>
<td>1220</td>
<td>2539</td>
</tr>
<tr>
<td>boat</td>
<td>350</td>
<td>1054</td>
<td>2304</td>
</tr>
<tr>
<td>bert</td>
<td>470–420</td>
<td>1270–1337</td>
<td>1540</td>
</tr>
<tr>
<td>boat</td>
<td>530–440</td>
<td>1088–900</td>
<td>2300</td>
</tr>
<tr>
<td>bout</td>
<td>684–420</td>
<td>1203–940</td>
<td>2538–2140</td>
</tr>
<tr>
<td>bite</td>
<td>660–400</td>
<td>1180–1880</td>
<td>2500–2524</td>
</tr>
<tr>
<td>bait</td>
<td>480–330</td>
<td>1720–2200</td>
<td>2520–2580</td>
</tr>
</tbody>
</table>

TABLE I. Center frequencies of F1, F2, and F3 for "bVt" stimuli.
normal. As a consequence, the extent of the categories was normal.

SS's identification of signals from the slit-split series is shown in Fig. 2. Once again, signals near the ends of the continuum were identified without variation and, in this instance, the category boundary fell just at the limit of normal performance.

SS's identification of signals from the VOT continua is shown in Fig. 3. Signals near the extremes of the continua were identified with consistency, and the phonetic boundaries were within the range of normal. SS's category boundary for the signals with high-F1 starting frequency fell at 24 ms, while the boundary for signals with low-F1 starting frequency fell at 32 ms. Both values were within the range of normal performance. The boundary shift (8 ms) was of similar magnitude to that of the normal listeners (6.5 ms).

2. Spectral continua

SS's identification of signals from the /i/--/e/ continuum is shown in Fig. 4. The signals near the extremes of the continuum were identified with near perfect consistency, and the category boundary fell at a normal value. The large difference in identification of stimulus 5 and 6 is particularly striking. Stimulus 5 was identified as /i/ 87.5% of the time, while stimulus 6 was identified as /e/ 100% of the time. The difference in the F1 and F2 frequencies of the two stimuli was only 20 Hz.

FIG. 1. Identification of signals from the chop-shop series. The dotted functions in this figure and in the following figures encompass the range of normal performance within ± 1 standard deviation of the mean. The solid function connecting the filled circles indicates SS's performance.

FIG. 2. Identification of signals from the slit-split series.

FIG. 3. Identification of signals from the VOT series with relatively high F1 starting frequency (top) and identification of signals with relatively low F1 starting frequency (bottom).

FIG. 4. Identification of signals from the /i/--/e/ series.
TABLE II. Percent "different" responses as a function of the difference in frequency between the "standard" stimulus at 400 Hz and "comparison" stimuli.

<table>
<thead>
<tr>
<th>Frequency difference (Hz)</th>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>80</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>120</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

One interpretation of this large change in identification with a small change in formant frequencies is that SS can resolve a small change in frequency of F1 (this follows from the view that F1 is the principal cue to vowel height). To obtain converging evidence on this point, we determined the difference limen for frequency of an isolated F1 resonance. The standard was a 400-Hz "F1," and the comparison stimuli were F1's of 420, 440, 460, 480, and 500 Hz. Each stimulus pair was presented ten times. The order of the pairs was randomized. The results shown in Table II. The point of 75% discrimination falls between 20 and 40 Hz.

SS's identification of the signals from the /a/-/e/ continuum is shown in Fig. 5. Identification of end point stimuli was poor, and performance, most generally, hovered around 50% identification.

SS's identification of signals from the save-shave continuum is shown in Fig. 6. Performance was within the range of normal. Thus SS can form categories that correspond to normal phonetic categories based on information that is relatively high frequency and voiceless (as in the save-shave stimuli) and based on information that is relatively low frequency and voiced (as in the /a/-/e/ stimuli).

SS's identification of signals that vary in onset frequency and direction of change of F2 and F3 (the ba-da series) is shown in Fig. 7. The stimuli at the da end of the continuum were identified at a level near chance. The stimuli at the ba end of the continuum were identified with slightly greater reliability. These outcomes were consistent with SS's verbal report of being unable to reliably identify the signals at the end points of the continuum in the familiarization phase of the experiment.

B. "bVt" stimuli

Identification of the bVt stimuli is shown in Table III. Identification accuracy in the full-cue condition averaged 99 percent correct for the normal hearing listeners and 79 percent correct for SS. Vowels synthesized with a relatively high F2; i.e., the front vowels /i e æ/ and the diphthongs /e ai/ were identified by SS with near perfect accuracy. Errors in identification were most common, then, for vowels with a relatively low F2. Inspection of the confusion matrix sug-
suggests that, most generally, error responses were vowels with similar F2 and/or F1. For Bert, the most common errors were bat and bet. The vowel /s/ is characterized by a distinctively low-frequency F3 at 1540 Hz. The vowels /ə/ and /ɛ/ have F2's at 1562 and 1610 Hz, respectively. It would appear that the error responses were conditioned by the similarity of the F2/F3 complex in Bert and F2's of bat and bet.

Identification accuracy in the F1-only condition averaged 39 percent correct for the normal listeners and averaged 33 percent correct for SS. Although the mean scores were similar, the stimuli were well identified differed greatly in several instances. For example, the normal listeners achieved much higher scores for /i/ (69%) and /ɪ/ (77%) than did SS (0% and 20%, respectively).

Identification accuracy in the F2 and higher formant condition averaged 69% correct for the normal listeners and averaged 41 percent correct for SS. SS's identification accuracy was poorer than normal for both vowels with relatively high F2 and for vowels with relatively low F2. The error responses for both the normal listeners and for SS are well accounted for by similarity in F2 of the target vowel and the error vowel.

C. Conflicting cue task

Identification of the bill, dill, gill stimuli is shown in Table IV. For normal-hearing listeners, cues that conflict with information from formant transitions change the identification of /b/ and /g/ very little. In the case of dill, identification accuracy improved when the duration of the steady-state spectrum was lengthened. Gill was not reported as a consequence of lengthening, which suggests that a lengthened onset was not a cue for "velar-ness." These data (and other data we have collected) suggest that, for normal-hearing listeners, formant transitions provide the most salient cue to place of articulation.

In contrast to the performance of the normal-hearing listeners, SS's performance was markedly affected by spectral shape at signal onset. Moreover, abruptness of frequency change following onset influenced identification responses in a different manner than for normal-hearing listeners. Spectral shape exerted a large effect on the identification of /b/ and /d/. For example, the stimulus with /b/ transitions and /b/ shape (i.e., a falling tilt and a low-frequency spectral peak) was heard as bill 88% of the time. However, the stimulus with /b/ transitions and /d/ shape (i.e., a rising tilt and a high-frequency spectral peak) was heard as dill 88% of the time. A similar, but smaller, effect is seen for stimuli with /d/ transitions. We suspect that for SS the location in frequency of the highest amplitude spectral peak at signal onset is a major cue to the /b/-/d/ distinction.

Identification of /g/, on the other hand, was greatly affected by abruptness of frequency change following onset. When the normally occurring, long steady-state following onset was removed from the gill stimuli, performance dropped from 100% to 50% gill responses. Given that the

duration of the burst and aspiration segment of the signal was perfectly correlated with abruptness of frequency change following onset, it is possible that SS was responding to a waveform, or envelope, cue rather than a cue in the frequency domain.

III. DISCUSSION

In this section, we first discuss the results of the identification tasks based on temporal cues. We then combine the results of the identification tasks based on spectral cues with the results of the bVt identification tasks and the conflicting cue task for discussions of the cues to vowel identity and stop consonant place of articulation. Finally, we comment on the relevance of our data for the design of cochlear prostheses.

A. Use of temporal cues

SS's identification of signals from the "temporal" continua was normal in all respects. The end points of the continua were reliably identified and the category boundaries were in normal locations. Thus signals that were phonetically unambiguous to normal listeners were unambiguous to SS, and signals that were phonetically ambiguous to normal listeners were ambiguous to SS. We infer from our results that, most generally, the implant transmits temporal information with sufficient resolution to allow the formation of categories that are isomorphic with the phonetic categories of stop consonant manner and stop consonant voicing established by normal listeners.

The implant also transmits information about F 1 onset frequency in a form that allows normal integration of low-frequency information and time information. We infer this from the shift in category boundary along the VOT continuum with a change in F 1 onset frequency. Indeed, the similarity in magnitude of the boundary shift for SS and for normal-hearing listeners suggests that the auditory representations of F 1 and of VOT elicited by the implant are very similar to those elicited by normal cochlear stimulation.

We note, parenthetically, that the location of the category boundary for the normal listeners and for SS in the low-F 1 condition was displaced away from the midpoint of the stimulus range. This outcome suggests that neither the normal listeners' nor SS's identification functions came about by a strategy of dividing the stimulus range into halves.

IV. VOWEL IDENTIFICATION

SS identified the full-cue bVt stimuli with impressive accuracy (mean = 79%). This level of performance suggests that information from both F 1 and F 2 was used in vowel identification. This suggestion was confirmed by the results of the F 1-only and F 2 and higher formants tests in which identification accuracy averaged 33% and 41%, respectively.

Our analysis of confusions in the bVt task indicated that the formant configuration that caused the most difficulty for SS was one in which F 2 was relatively low. This outcome prompted the development of the /s/-/z/ continuum in which F 1 was fixed and F 2 varied over the range 880–1180

TABLE IV. Mean percent response as a function of stimulus factors and hearing status. Tilt refers to spectral shape at signal onset. Abruptness refers to the duration the spectrum stays steady state, or nearly so, following onset [abrupt (a) change = steady state of approximately 9 ms; nonabrupt (na) change = steady state of approximately 27 ms].

<table>
<thead>
<tr>
<th>Stimulus factors</th>
<th>Normal</th>
<th>Implant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitions</td>
<td>Tilt</td>
<td>Abruptness</td>
</tr>
<tr>
<td>b b a</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>b d a</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>b g a</td>
<td>94</td>
<td>1</td>
</tr>
<tr>
<td>b b na</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>b d na</td>
<td>93</td>
<td>3</td>
</tr>
<tr>
<td>b g na</td>
<td>91</td>
<td>1</td>
</tr>
<tr>
<td>d d a</td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td>d b a</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>d g a</td>
<td>25</td>
<td>74</td>
</tr>
<tr>
<td>d d na</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>d b na</td>
<td>10</td>
<td>83</td>
</tr>
<tr>
<td>d g na</td>
<td>12</td>
<td>82</td>
</tr>
<tr>
<td>g g a</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>g b a</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>g d a</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>g g na</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>g b na</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>g d na</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
Hz. As we suspected, SS was unable to reliably identify the end points of the continuum.

In contrast, SS identified signals in a normal manner from the continuum (/a/-/e/) in which F1 and F2 covaried. As we noted in Sec. II, performance in this task was characterized by a large change in identification with a 20-Hz change in F1 and F2. This outcome prompted the test for the difference limen for F1. The result of that test, a difference limen of between 20 and 40 Hz, was consistent with the outcome of the identification task. Taken together, the results indicate that SS can resolve and use in a phonetically useful manner small differences in F1 frequency. Certainly, resolution is sufficient to resolve the F1 differences of English vowels.

Recall that SS identified front vowels and the diphthongs /e/ and /ai/ with near perfect accuracy. This outcome, and the outcome of a normal difference limen for F1, led us to ask: If the location of F1 is accurately resolved, how accurate must be resolution of F2/F3 in order to achieve the level of front vowel recognition achieved by SS?

The outcome of a recent experiment with normal-hearing listeners by Dubno and Dorman (1987) bears on this issue. Synthetic tokens of beet, bet, bat, bait, and bite were modified in two ways. For one set of signals, energy above F1 was removed by digital filtering. For a second set, the spectral peaks of F2-F3 were flattened by maximally broadening the bandwidths of F2-F3 in the synthesis parameter files and by additional digital filtering. The resulting stimuli had normal F1's and a broad, flat plateau from approximately 1600 to 2500 Hz. The F1-only stimuli were identified with a mean accuracy of 33%. The stimuli with normal F1 and a high-frequency plateau were recognized with 96% accuracy. We interpret this outcome to mean that, if F1 is well resolved, then poor resolution of the location of higher formants is sufficient for normal identification of front vowels. It is possible that the auditory information available to SS is like this, i.e., an accurate representation of F1, given by the fine structure of the time waveform, and/or by the distribution of activity in electrodes 1 and 2, and a less accurate indication of the frequency of higher formants given by the distribution of activity in electrodes 2-4.

There is, however, another way to account for accurate identification of front vowels. Two of the normal-hearing listeners in Dubno and Dorman (1987) achieved a mean score of 93% correct for the F1-only stimuli. Thus information from F1 can be a sufficient condition for identification of front vowels. While this was not the case for SS, for whom front-vowel (and diphthong) recognition in the F1-only condition averaged 30 percent correct, it may contribute in significant measure to the vowel recognition scores of some users of cochlear implants, especially those using single-channel devices (see, for example, White, 1983).

A. Stop consonant recognition

The results for SS from the /ba/-/da/ continuum suggest that the information provided by formant transitions of relatively brief duration is of little use in stop consonant recognition. This suggestion is bolstered by the results of the conflicting cue experiment, which indicated that, when formant information conflicts with onset spectrum information, the latter dominates identification of /b/, and, to a lesser degree, /d/. Together, the data indicate that SS bases the /b/-/d/ distinction more on the spectrum at onset than on the following formant transitions.

We can rationalize accurate identification of /b/ and /d/ based on onset cues by noting that the center frequencies of electrodes 2, 3, and 4 are at 1, 2, and 4 kHz, respectively. The relatively low-frequency energy concentration at the onset of /b/ will have the greatest effect at electrode 2. The higher frequency energy at the onset of /d/ will have the greatest effect at electrodes 3 and 4. If electrodes 2, on the one hand, and electrodes 3 and 4, on the other, produce discriminable outputs, then b's and d's ought to be identifiable.

The results of the conflicting cue experiment suggest that the /d/-/g/ distinction is based on the relatively long spectral steady state following the onset of /g/. This we infer from the 50% decrease in g responses when the long steady state was removed from gill and from the tendency to respond g when the /b/ and /d/ onsets were lengthened to a duration appropriate for /g/. The use of a temporal cue for the /d/-/g/ distinction is not surprising given that the location of the spectral peaks for /d/ and /g/ in the /b/ environment differ only by 600 Hz. That is to say, the use of temporal cue is reasonable given poor frequency resolution in the region of the spectral peaks. (We note that this outcome is not novel, but rather is also seen in hearing-impaired listeners (Lindholm et al., 1988).)

B. On the design of cochlear implants

Designers of cochlear prostheses must have answers to, or at least opinions on, the following issues: Should the speech signal be presented in analog form, or, should features, such as formants, be extracted before presentation to a patient's auditory system? How many channels of stimulation are necessary? Should the electrodes be monopolar or bipolar? (Here, the issue is whether the spread of current from monopolar stimulation will eliminate information about place of stimulation.) SS’s success in phonetic identification [and the success of other patients using the Symbiont Prosthesis (Gantz et al., 1987)] demonstrates that a speech processing system that drives a small number of monopolar electrodes with bandlimited, analog waveforms is a viable option for the design of a cochlear prosthesis.

ACKNOWLEDGMENTS

We thank Korine Dankowski for assistance and thank SS for his patience and insights during the many hours of testing. This research was supported by a grant (RO1 NS 15457-6) from the National Institutes of Health and by grants from the College of Liberal Arts and Sciences, Arizona State University.