

An AER study of stop-consonant discrimination

MARYLOU PAUSEWANG GELFER
Indiana University, Bloomington, Indiana

The purpose of this study was to explore hemispheric involvement in stop-consonant discrimination. Two experimental designs were used. In the first design, averaged evoked responses (AERs) to stop-consonant-vowel (CV) syllables were combined with AERs to nonspeech stimuli, in a paradigm similar to earlier studies, and were submitted to a principal components analysis and analyses of variance. In the second design, only the CV-syllable AERs were analyzed, in the same manner. When the responses to both CV and nonspeech stimuli were included in the analysis, the results were in agreement with those of earlier studies. However, when the nonspeech-stimuli AERs were removed from the analysis, the unilateral effects observed in prior studies were not replicated. The results of this research indicate the importance of considering experimental design and task variables before generalizing AER results to speech perception.

It has been assumed since the late 1800s that the left hemisphere of the brain is somehow specialized for language. Indeed, the pervasiveness of various types of aphasia following injury to or disease of the left hemisphere in right-handed individuals gives credence to this view. But what exactly is the left hemisphere's role in speech perception? This question has not been an easy one to investigate. "Speech" is composed of many acoustically diverse elements, and hemispheric involvement in processing these elements has been difficult to determine.

Results of dichotic listening studies have revealed a right-ear advantage (REA)—corresponding to a presumed left-hemisphere superiority—for certain types of phonetic stimuli (Kimura, 1961). Shankweiler and Suddert-Kennedy (1967) and Cutting (1974) found that an REA existed for stop-consonant-vowel (CV) stimuli. Stop consonants elicited the most marked REA, whereas the liquids /r/ and /l/ elicited a smaller REA (Cutting, 1974). Steady-state vowels did not appear to elicit a significant REA in either study.

Molfese (1978) attempted to replicate some of Cutting's results but employed averaged evoked responses (AERs) rather than dichotic listening to demonstrate the differential hemispheric responses. Neuroelectric activity was measured by electrodes at T3 and T4 of the 10-20 electrode system (Jasper, 1958), referenced to linked earlobes. In a paradigm similar to Cutting's (1974) investigation, Molfese used CV syllables with normal (phonetic) transitions, CV syllables with inverted (nonphonetic) transitions, and a bandwidth variable—sine-wave-formant CV analogs—which had both phonetic and nonphonetic transitions. It should be noted that these stimuli, particularly those with sine-wave formants, are not readily compre-

hended as spoken syllables, although with training they can be perceived as "computer-generated speech" (Schwab, 1981). Their inclusion within the set of stimuli to be discriminated could be expected to significantly increase the difficulty of the discrimination, thereby introducing a task variable.

Molfese (1978) used principal-components analysis in analyzing AER data. This procedure permitted him to identify underlying components of the AERs that might vary with experimental manipulation. Results revealed that both /b/-/g/ with phonetic transitions and /b/-/g/ with nonphonetic transitions were differentiated in the left hemisphere, but in different ways. Thus, the left hemisphere appeared to be sensitive to normal-transition /b/-/g/ contrasts, nonphonetic /b/-/g/ contrasts, and normal versus nonphonetic transitions. No such differences were observed in the right hemisphere. Importantly, however, the bandwidth variable was not a significant factor in this interaction. That is, in assessing left hemisphere sensitivity to /b/-/g/ contrasts, the responses to normal-bandwidth syllables and sine-wave-bandwidth CV analogs were not differentiated, and, in addition to the previously mentioned differences in task difficulty, some of the sine-wave-formant syllables could be expected to be mislabeled. These (presumed) incorrect perceptual responses were not removed from the analysis. Thus, in this study, the confounding variables of stimulus difficulty and incorrect responses were not addressed.

Molfese (1980) added a dimension of varying vowel contexts to the paradigm described in his earlier study. In this research the /b, g/ consonants were randomly paired with the vowels /i, æ, ɔ/. Subjects again heard both normal-bandwidth and sine-wave-analog CVs, although nonphonetic (inverted) transitions were not utilized. Electrode sites in this study included T3, T5, and P3 over the left hemisphere, and T4, T6, and P4 over the right hemisphere (Jasper, 1958), referenced to linked earlobes. Results of this research revealed that adult subjects tended to discriminate both normal-bandwidth and sine-wave-

This research was based on the author's dissertation under the direction of Harry Hollien at the University of Florida. The author gratefully acknowledges his support of this project. The author's address is Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405.

formant /b/ and /g/ in an initial bilateral cortical process with a peak latency of 170 msec. A later unilateral process was observed with latencies at 215, 290, and 460 msec post stimulus.

Molfese and Schmidt (1983) essentially replicated this procedure and obtained similar findings. In both studies, results showed a significant hemisphere \times consonant interaction such that /b/ and /g/ (regardless of vowel environment or bandwidth) were differentiated in the left hemisphere but not in the right. Molfese (1980) and Molfese and Schmidt (1983) were the first AER studies to reveal a consistent left-hemisphere response to consonants in varying vowel contexts. This was a significant finding, because the acoustic cues for each consonant vary with the following vowel (see, e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

However, in these studies, the effects of several possible confounding factors were not examined. First, as discussed above, one set of stimuli (the sine-wave-formant CV analogs) was less familiar to the subjects than the other, and more difficult to discriminate; it is possible that a listening task that involved discrimination between these elements would have required a greater degree of attention from the subjects. Second, some of the subjects' perceptual judgments of the ambiguous stimuli (sine-wave-formant CV analogs) would almost certainly have been incorrect. Finally, like most other research in this area, the Molfese studies utilized synthetic syllables consisting of initial-formant transitions and vowel formants only. Although it is usually assumed that such stimuli will be perceived in the same manner as natural speech, this assumption has rarely been tested.

Regarding the differences in stimulus difficulty and required attention, the results of numerous AER studies suggest that as difficulty in discriminating among stimuli increases, AER latencies become longer (Ritter, Simson, & Vaughn, 1972) and amplitude increases (Poon, Thompson, & Marsh, 1976). Other studies have shown that as the amount of attention required by a task increases, so do AER amplitudes (Eason, Harter, & White, 1969; Harter & Salmon, 1972). Furthermore, dichotic-listening studies have shown that increasing task difficulty results in larger hemispheric differences. For example, when listeners were asked to identify vowels in noise (Weiss & House, 1973) and vowels of brief duration (Godfrey, 1974), a tendency toward REA increased. Furthermore, Kasischke (1979) demonstrated that increasing the complexity of tonal stimuli resulted in asymmetric hemispheric involvement. Thus, it is possible that the left-hemisphere /b/-/g/ discrimination found in the Molfese research was dependent upon the inclusion of ambiguous stimuli in the research design, and does not reflect normal speech perception.

The second confounding variable mentioned above, incorrect perceptual judgments, is also potentially serious. If one assumes that electrocortical activity reflects a cognitive process or series of processes, an incorrect perceptual judgment should result in a slightly different wave-shape than a correct perceptual judgment. Thus, it would

appear important to include only correct perceptual judgments when averaging trials to obtain AERs.

Finally, when synthetic syllables are utilized in perceptual experiments, differences in formant-transition onset frequency are often used as the only cue differentiating one consonant from the other. However, it has been demonstrated by Kewley-Port (1982) that formant transitions alone are not sufficient cues in natural speech for accurate stop-consonant identification, despite their frequent use in speech perceptual studies. It is possible that perceptual processes identified generally in the literature could vary significantly as a function of the type of speech stimulus used (synthetic vs. natural).

In summary, uncontrolled task or stimulus variables could have affected the results obtained in previous AER studies. A research design that takes into account stimulus difficulty/required attention and accuracy of judgments would appear to be necessary to separate hemispheric response to stimulus characteristics from hemispheric response to task variables.

The purpose of the present experiment was to further examine hemispheric involvement during stop-consonant discrimination. Specific questions addressed include the following: (1) Are the consonants /b/ and /d/ differentiated in the left hemisphere, not the right, when responses to both speech-like syllables and nonspeech analogs are averaged, in a paradigm similar to Molfese (1980) and Molfese and Schmidt (1983)? (2) Are /b/ and /d/ differentiated in the left hemisphere, not the right, when task difficulty is reduced (i.e., only correct responses to speech-like syllables are used)? (3) Do synthetic speech stimuli and natural speech stimuli elicit evoked responses that differ significantly in terms of latency and hemispheric symmetry/asymmetry during perception of /b/ and /d/ syllables?

METHOD

Stimuli

The stimuli for this study included the synthetic and natural syllables /bi, bæ, bɔ, di, dæ, dɔ/. Also included were the "chirps" associated with these syllables, that is, their isolated second- and third-formant transitions; with the fundamental frequency and first, fourth, and fifth formants eliminated. See Figure 1 for a representation of these stimuli. The chirps taken from the six syllables sounded somewhat like tongue clicks. Although they were perceptually different from each other, they were not readily recognizable as speech, and were included as the nonspeech analogs for this experiment.

The synthetic syllables consisted of a 50-msec transition followed by a 300-msec steady-state vowel segment. Specific onset values for each transition and their associated steady-state vowel formants are shown in Table 1. Transition onset frequencies were taken from data presented by Kewley-Port (1982) and Klatt (1980), and were modified as necessary during the synthesis to achieve optimal discriminability. Vowel-formant frequencies for F1 through F3 were taken from Peterson and Barney's (1952) data. It will be noted that F4 and F5 are constant across the entire syllable duration, and are the same for each vowel. The upper formants were included to make the synthetic syllables sound more natural.

For all syllables, the bandwidth for F1 was 60 Hz; for F2, 90 Hz; and for F3 through F5, 120 Hz (Cutting, 1974; Molfese, 1978, 1980; Molfese & Schmidt, 1983). Each syllable had an associated

Synthetic Syllables

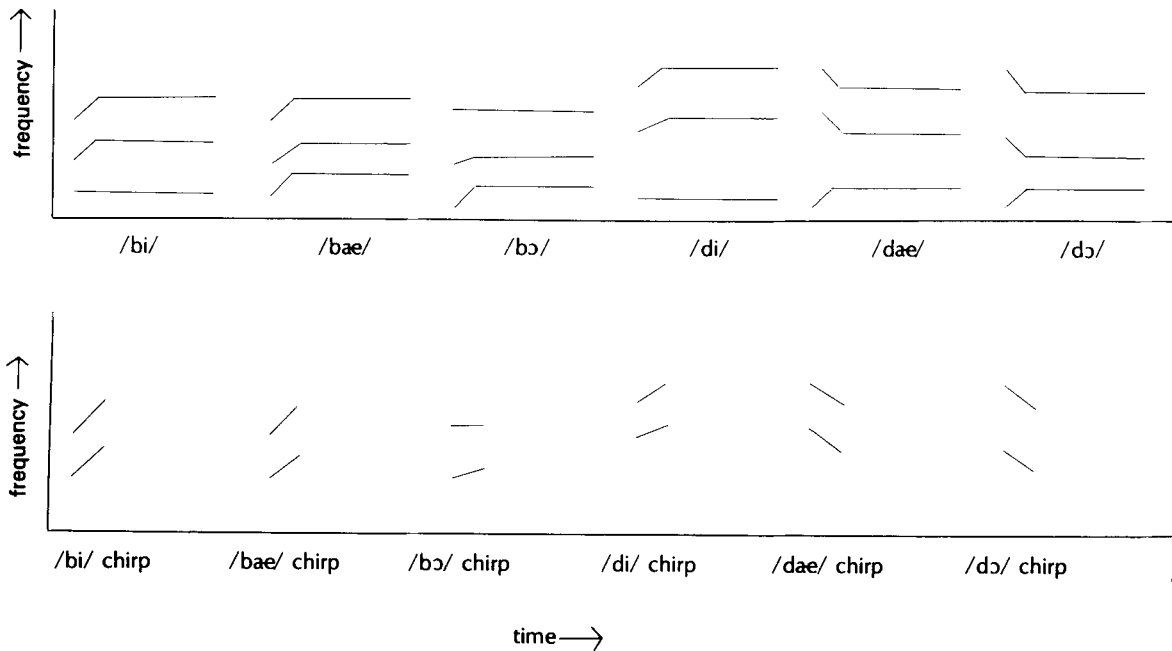


Figure 1. Schematic representation of the synthetic syllable stimuli and chirp stimuli.

fundamental frequency of 130 Hz (Peterson & Barney, 1952), and a rise time of 30 msec. All synthetic syllables were produced by a Klatt software synthesizer (Klatt, 1980) and recorded on one channel of a TEAC 6120 dual-channel tape recorder.

The chirp stimuli were taken from their respective synthesized complete syllables. When the program for each syllable was run, only the first 50 msec (the transition portion) was activated. A digital filter was utilized to eliminate fundamental frequency, F1, F4, and F5.

The natural-syllable stimuli were produced by a male speaker who had clearly identifiable vowel formants and the ability to modify fundamental frequency upon request. During production of the stimuli, the speaker was seated in a double-walled Industrial Acoustics Company (IAC) booth. The stimuli were recorded using a B&K 5065 half-inch condenser microphone and a B&K 37A preamplifier, coupled with a Revox B-77 tape recorder. The best examples of each syllable were modified by eliminating prevoicing of the consonant (but retaining the spike) and by reducing vowel duration to conform as closely as possible to the synthetic stimuli (50 msec transition duration, 300 msec vowel duration). Mean transition duration of the selected syllables was calculated to be 48 msec (range: 24-72 msec). Mean vowel duration was found to be 292 msec (range: 262-304 msec). Finally, the rise time of each syllable was calculated from the output of a Honeywell 1508A Visicorder, and the exemplars with a rise time most closely approximating 30 msec were selected for inclusion as natural-syllable stimuli. Mean rise time of the selected syllables was 38.1 msec (range: 31.2-41.6 msec). Actual onset (transition) and steady-state (vowel) frequencies for the first three formants of the natural-syllable stimuli are provided in Table 2.

Stimulus Tape Construction

Three stimulus tapes, one each with synthetic syllables, with natural syllables, and with chirps, were constructed for presenta-

Table 1
Onset and Steady-State Frequencies for Synthetic Syllables

Syllable	Formant	Onset Frequency (Hz)	Steady-State Frequency (Hz)
/bi/	F1	200	310
	F2	1100	2020
	F3	2150	2960
	F4	3300	3300
	F5	3750	3750
/bae/	F1	200	620
	F2	1100	1660
	F3	2150	2430
	F4	3300	3300
	F5	3750	3750
/bo/	F1	200	600
	F2	900	990
	F3	1900	2570
	F4	3300	3300
	F5	3750	3750
/di/	F1	200	310
	F2	1800	2020
	F3	2960	2960
	F4	3300	3300
	F5	3750	3750
/dae/	F1	200	620
	F2	1600	1660
	F3	2700	2430
	F4	3300	3300
	F5	3750	3750
/do/	F1	200	600
	F2	1600	990
	F3	2700	2570
	F4	3300	3300
	F5	3750	3750

Table 2
Onset and Steady-State Frequencies for Natural Syllables for the First Three Formants

Syllable	Formant	Onset Frequency (Hz)	Steady-State Frequency (Hz)
/bi/	F1	250	313
	F2	1800	2000
	F3	2375	2850
/bæ/	F1	625	750
	F2	1500	1625
	F3	2313	2450
/bo/	F1	500	625
	F2	938	1000
	F3	2225	2225
/di/	F1	188	313
	F2	1937	2125
	F3	2765	3000
/dæ/	F1	380	750
	F2	1700	1375
	F3	2650	2375
/do/	F1	385	610
	F2	1500	1000
	F3	2480	2225

tion during the experimental session. Each of the six syllable (or chirp) stimuli was repeated 20 times in random order on each tape, for a total of 120 stimuli per tape. The 120 stimuli were recorded on both channels of an Akai GX-77 tape recorder. Interstimulus intervals were varied from 2 to 9 sec to avoid eliciting a time-locked cortical expectancy response. Maximum amplitude of each syllable or chirp was monitored on the VU meter of the Akai tape recorder, and adjusted prior to recording so that all stimuli peaked at 0 VU.

Subjects

The subjects were 12 young adults—6 males and 6 females—aged 23–33 years (mean age = 28.0 years). All subjects were majoring in or employed in the fields of experimental phonetics or speech pathology. In the first selection protocol, the subjects were required to demonstrate pure-tone thresholds of better than 20 dB at 0.5, 1.0, 2.0, 4.0, and 8.0 kHz, with a mean between-ear threshold difference of less than 5 dB. In addition, any potential subject with a 10-dB or greater between-ear threshold difference at any single frequency was rejected. Second, subjects were administered the Edinburgh Handedness Inventory, or EHI (Oldfield, 1971), and selected on the basis of a strong right-hand preference. The average laterality quotient for all subjects on the EHI was 92.8 out of a possible 100.0 (range = 83–100), with a mean decile of 8.2 out of possible 10.0 (range = 6–10). As a final step in the selection process, a screening test of the synthetic stimuli was presented. The subjects were asked to listen to a sequence of 60 of the synthetic CV syllables used in the experimental sessions. The subjects then identified the initial consonant they heard for each syllable. A score of 95% correct or better on the 60-item screening test was required for volunteer subjects to be included in the experiment. The subjects were allowed up to three attempts to attain a score of 95%. On the final trial, mean percent correct was 98.5%, with a range of 97–100%.

Procedure

Following successful completion of the selection protocols, chlorided silver electrodes (Grass E6SH) were placed on each subject's scalp in the superior temporal regions of both hemispheres (T3 and T4, as described in Jasper's 10-20 electrode system, 1958). These active sites were referenced to contralateral earlobes (G. Goff, Matsumiya, Allison, & W. Goff, 1977; W. Goff, 1974; W. Goff, Mat-

sumiya, Allison, & G. Goff, 1969).¹ Additionally, one electrode was placed above the inside corner of each subject's right eye, and another was placed at the lateral superior aspect of the right orbital ridge, to record extraocular eye movements and blinks for an artifact-rejection channel. Finally, one ground electrode was placed on the left mastoid process. Electrode resistances between T3 and the right earlobe averaged 3.82 k Ω and between T4 and the left earlobe averaged 4.03 k Ω . Final resistances measured at the completion of the experimental session averaged 4.39 and 4.59 k Ω for the left and right active sites, respectively.

During the electrocortical recording protocols, the subjects reclined on a bed in a double-walled electrically shielded IAC booth. They were instructed to keep their eyes closed, jaws relaxed, and to move as little as possible when signaling their perception of /b/ or /d/ (by raising both index fingers or both fourth fingers). The syllables or chirps recorded on the right channel of an Akai GX-77 tape recorder were played through a Kenwood KA-7100 amplifier outside the IAC booth, to an ADS 810 speaker located inside the booth, at an intensity level of 62 dB re: .0002 dynes/cm² at the subjects' ear. The speaker was positioned approximately 78 in. directly in front of the subject. The stimuli recorded on the left channel of the tape recorder were input directly to a Schmitt trigger, which was utilized to synchronize stimulus onsets during the averaging procedure.

Cortical responses from the two active sites (T3 and T4) were amplified by two Grass 7P122A amplifiers with bandpass flat from .04 to 60 Hz. System gain was set at 28,000. The electrodes around the eye were connected to a Grass 7DAF DC Driver Amplifier and 7P3B Preamplifier, with bandpass flat from .3 to 75 Hz and system gain set at 11,000. Output from the three amplifiers was input into an analog-to-digital (A/D) conversion device and a Digital Equipment Corporation PDP 11/23 computer. The Schmitt trigger pulse was fed into a fourth channel of the A/D board. At the occurrence of each pulse on Channel 4, corresponding to the onset of each stimulus item, the electrocortical waves input into Channels 1 through 3 were digitized at 200 Hz for 500 msec. These digitized waves were stored on hard disk for later selective averaging on the PDP 11/23, that is, the experimenter was free to exclude any single response from the averaging process, so that, if desired, only electrocortical responses associated with correct perceptual judgments could be utilized. Activity recorded from the periorcular site that exceeded a preset threshold (determined for each subject in a pretest) caused data from T3 and T4 to be automatically excluded from the averaging process.

The subjects were initially presented with one stimulus tape containing exclusively synthetic syllables, then another containing exclusively natural. Order of presentation was randomized and balanced over subjects. Following the two syllable tapes, the subjects were given a short training session to familiarize them with the chirp stimuli. During this training, half of the subjects were instructed to perceive the chirps as speech and to discriminate the /b/ and /d/ chirps. The other half of the subjects were instructed to perceive the chirps as frequency glides, and to discriminate high-versus low-frequency onsets. After listening to the stimulus tape, each group was then given the alternate instructions and training, and the chirp stimulus tape was presented again. Order of presentation (speech instructions, frequency instructions) was randomized and balanced over subjects. Thus, the three 120-item stimulus tapes were presented as four trial blocks: synthetic syllables, natural syllables, chirps with speech instructions, and chirps with frequency instructions.²

RESULTS

Perceptual Results

During the electrocortical recording procedure, the subjects responded to the synthetic syllables with 97.8% ac-

curacy (range: 93%–100%) and to the natural syllables with 98.9% accuracy (range: 96%–100%). A *t* test for related measures revealed no significant differences ($\alpha > .05$) between these means, indicating no differences in accuracy between subjects' responses to synthetic and natural syllables. When individual electrocortical responses were averaged in calculating each subject's AER for each syllable, responses associated with incorrect perceptual judgments and muscle artifacts were excluded. This resulted in utilizing an average of 19.5 responses per syllable for the synthetic-syllable AERs and 18.7 responses per syllable for the natural-syllable AERs.

For the chirp trial (speech instructions), the subjects were able to discriminate /b/ from /d/ in the three vowel contexts with 57.2% accuracy (range: 24%–83%). Because of the low percentages of correct responses, elimination of all incorrect responses was not possible. As a result, all chirp responses not contaminated by muscle artifacts were included in the averaging process, for an average of 19.7 responses per chirp utilized in obtaining each chirp AER.

Principal-Components Analysis 1

The electrocortical responses were selectively averaged, as described above, for each of the 12 subjects, 6 syllables, 2 hemispheres, and 4 trials, and the resulting 576 AERs were normalized. The entire data set was then submitted for principal-components analysis (PCA) and analyses of variance (ANOVAs) (Chapman, McCrary, Bragdon, & Chapman, 1979; Molfese, 1978, 1980, 1984; Molfese & Schmidt, 1983).

The first step in the PCA was to calculate the centroid, or average, of all 576 normalized AERs (Dixon, 1981). The centroid is pictured in Figure 2. It is characterized by a small positive peak at 45 msec (P45), a large negative peak at 120 msec (N120), a large positive peak at 195 msec (P195), a negative peak at 270 msec (N270), a small positive peak at 340 msec (P340), followed by a gradual negative decline that asymptotes at 455 msec

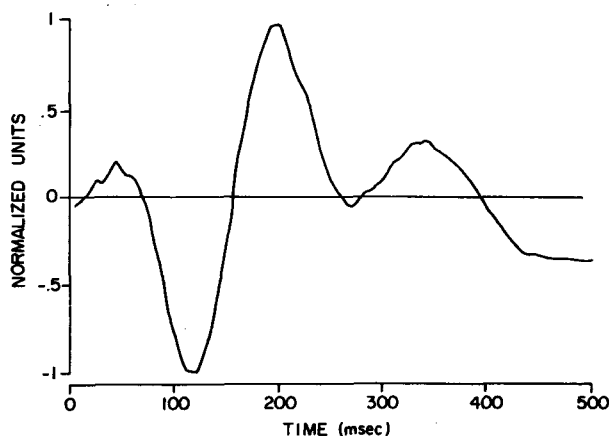


Figure 2. Centroid for Analysis 1, based on the full data set of 576 AERs.

(N455). This centroid is very similar in waveshape to the one reported by Molfese and Schmidt (1983), which showed a waveshape of P30, N120, P200, N270, P345, and N450. The main difference occurred in the final 150 msec, during which the present experiment found a falling configuration, whereas Molfese and Schmidt (1983) found a level to rising configuration.

The next step in the PCA was the formation of a 100×100 covariance matrix and the extraction of the principal components (or factors). Factors with eigenvalues of one or more were retained for further analysis (Chapman et al., 1979). This criterion insured that a factor was not retained unless it accounted for at least as much variance in the data as any one of the original 100 variables. This procedure resulted in 10 factors, which accounted for 62.7% of the variance observed in the data. The factors were then rotated using a varimax criterion (Kaiser, 1958) to maintain orthogonality while improving factor distinctiveness. After 14 iterations, the terminal solution was reached. The rotated factors are pictured in Figure 3. The peaks observed in these plotted factors, or component waves, indicate latency regions in which critical differences occur between AERs in different hemispheres to various stimuli. These regions are assumed to be of importance, to a greater or lesser extent, in each individual AER. For example, if the peak latency of a factor were 100 msec, some percentage of AERs in the data set would be expected to vary in the region immediately surrounding 100 msec.

The final step in the PCA was the calculation of 10 sets of factor scores for each of the original 576 AERs (based on the 10 extracted factors). Thus, each AER in the data set was effectively represented by 10 factor scores in the place of its original 100 voltage values.

At this point, factor scores for each AER were utilized as the dependent variables in 10 ANOVAs (one for each factor). All possible main effects and interactions for the independent (classification) variables of consonant, vowel, hemisphere, and trial were included in the statistical model (Dixon, 1981). Because the factors were specifically calculated to be orthogonal, independent ANOVAs could be appropriately performed.

In assessing the significance of ANOVA results, a probability level of $\alpha = .05$ was chosen in order to include as many main effects and interactions as possible while maintaining a reasonably high level of significance.

PCA and ANOVA Results for Analysis 1

The principal question addressed in Analysis 1 was whether /b/ and /d/ were differentiated in the left hemisphere for trials that included both syllable stimuli (synthetic and natural) and ambiguous stimuli (chirps with speech instructions). This paradigm was somewhat similar to that used by Molfese (1980) and Molfese and Schmidt (1983).

To test this relationship, the 10 ANOVAs described above were examined for significant consonant \times hemisphere \times trial interactions. The ANOVA of one factor

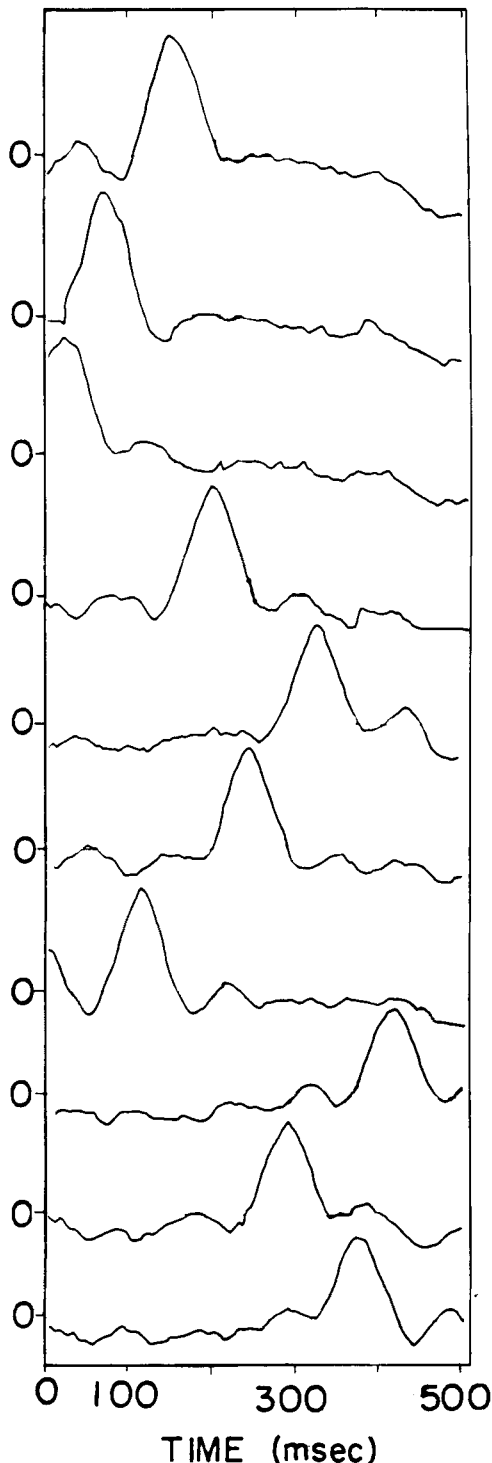


Figure 3. Factors 1 through 10 resulting from a principal-components analysis of the full data set.

(Factor 9) did indeed reveal a significant consonant \times trial \times hemisphere interaction [$F(3,33) = 3.63, p = .0229$]. An a priori test using an F statistic (Kirk, 1982) was carried out, in which left-hemisphere mean factor scores for /b/ in synthetic syllables, natural syllables, and chirps with

speech instructions, were compared to those associated with /d/. Attained significance for this comparison was $P = .0118$ [$F(1,33) = 7.02$]. This comparison clearly replicated the major findings reported in the Molfese research (Molfese, 1978, 1980; Molfese & Schmidt, 1983), in spite of the differences in stimuli and other experimental factors. Stop-consonant-syllable stimuli, averaged with CV analogs (in this case chirps rather than sine-wave-bandwidth formants) appeared to be differentiated in the left hemisphere. As in the Molfese research, a comparison between consonants in the right hemisphere revealed chance-level discrimination [$F(1,33) = .204; p = .659$]. Finally, a comparison of the factors associated with left-hemisphere discrimination in Molfese (1980), Molfese and Schmidt (1983), and this experiment showed some minor but interesting similarities. In Molfese (1980) and Molfese and Schmidt (1983), the major peak latency of such factors occurred at 460 msec, with smaller peaks occurring at 215 and 290 msec. In this research, the major peak latency of the factor associated with left-hemisphere discrimination was at 290 msec, with minor peaks occurring at 385 and 450 msec. Although the factor waveshapes among studies were quite different, the locations of the major and minor peaks (290 and 450–460 msec) suggest a generally similar time course for these component waves.

In addition to left-hemisphere unilateral processing of stop consonants, Molfese (1980) and Molfese and Schmidt (1983) found a bilateral processing stage, during which /b/ and /g/ appeared to be differentiated regardless of hemisphere. The principal latency of the factor associated with this discrimination was 170 msec. In the present research, a main effect for consonant discrimination independent of hemisphere was also revealed in the ANOVA results for Factor 1 [$F(1,11) = 15.98, p = .0021$]. The main peak characterizing Factor 1 occurred at a latency of 150 msec. In this case, the results of previous research and the findings of this experiment appear to be in agreement, for both the significance of the main effect and the location of the peak latency.

PCA and ANOVA Results for Analysis 2

To answer the second research question posed in the introduction, a separate PCA and series of ANOVAs were calculated for the syllable data only. Such analyses were done to exclude AERs associated with a difficult task and frequent incorrect responses (43% of the chirp responses included in Analysis 1 were erroneous).

Input to the PCA of Analysis 2 consisted of 288 normalized AERs based on 12 subjects \times 2 consonants (/b, d/) \times 3 vowels (/i, æ, ɔ/) \times 2 hemispheres \times 2 trials (synthetic syllables and natural syllables). The centroid is pictured in Figure 4. It is characterized by a small positive peak at 40 msec (P40), a large negative peak at 115 msec (N115), a large positive peak at 190 msec (P190), a negative peak at 265 msec (N265), a positive peak at 330 msec (P330), followed by a gradual decline that asymptotes at 490 msec (N490). This centroid is very similar to the one obtained from the full data set utilized in the first analysis.

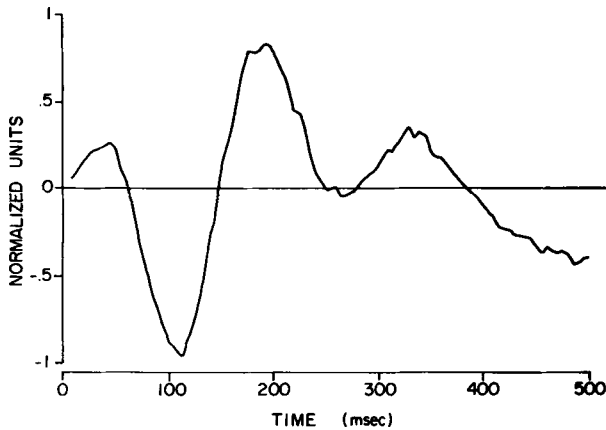


Figure 4. Centroid for Analysis 2, based on responses to synthetic and natural syllables only (288 AERs).

The PCA was calculated in a manner similar to that described for the full data set. Ten factors were extracted, which accounted for 79.7% of the observed variance in the data. The rotated factors are pictured in Figure 5. Factor scores were then calculated and submitted to an ANOVA program (Dixon, 1981). This procedure resulted in 10 new ANOVAs based on the 10 extracted factors, each of which evaluated all possible main effects and interactions among the independent variables. As in the previous analysis, left-hemisphere differentiation of /b/ and /d/ was the issue of primary interest.

Based on the results of Molfese (1980, 1983), and Analysis 1 of this experiment, it was hypothesized that /b/ and /d/ would be differentiated in the left hemisphere, not the right. To test this hypothesis, a consonant \times hemisphere interaction was sought in the 10 ANOVAs based on the 10 extracted factors. The results revealed no such interaction significant at the $p = .05$ level or better. Thus, when only synthetic and natural syllables were analyzed, a left-hemisphere differentiation between /b/ and /d/ was not obtained.

Despite the lack of a unilateral processing stage, a main effect for consonant indicated that the previously observed bilateral response was again obtained. The ANOVA on Factor 3 revealed a main effect for consonant at the $p = .0145$ level [$F(1,11) = 8.41$]. This result indicated that /b/ and /d/ were again differentiated bilaterally. Furthermore, the latency of this bilateral process (factor-waveshape peak at 160 msec) was in agreement with the latency determined in the Molfese research (170 msec) and in Analysis 1 (150 msec). For this effect, elimination of the chirp trials did not appear to alter the obtained pattern of significance and latency.

The final research question posed in the introduction involved a comparison of the AERs elicited by synthetic versus natural syllables to determine if conclusions based on synthetic syllables could be generalized to natural syllables. This question was addressed in two ways. First, the 10 ANOVAs (based on the 10 factors isolated in Anal-

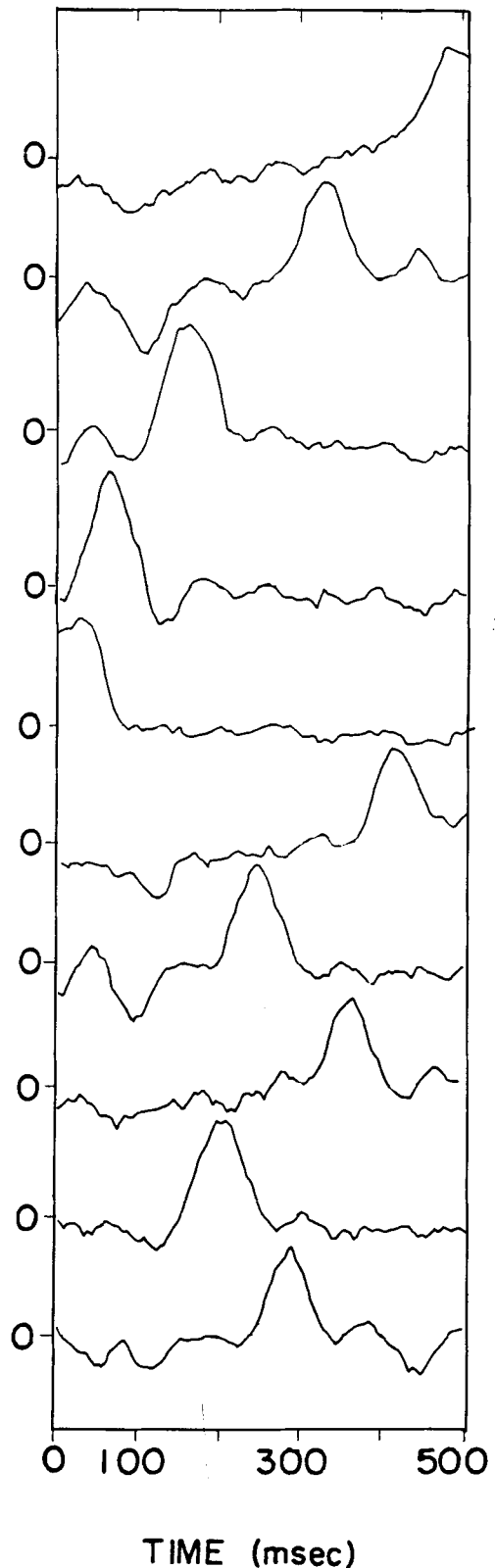


Figure 5. Factors 1 through 10 resulting from a principal-components analysis of responses to synthetic and natural syllables.

ysis 2) were scanned for main effects for the trial variable. The results revealed that a main effect for trial characterized Factors 2, 3, 4, and 6 [$F(1,11) = 15.71$, $p = .0022$; $F(1,11) = 15.63$, $p = .0023$; $F(1,11) = 26.98$, $p = .0003$; $F(1,11) = 7.53$, $p = .0191$; respectively]. Since synthetic syllables and natural syllables were the only levels included in the trials variable in Analysis 2, synthetic syllables appeared to elicit significantly different AERs than did natural syllables. However, the effect of this difference on the variables in question—unilateral and bilateral /b/-versus-/d/ discrimination—needed further investigation. The 10 ANOVAs, therefore, were again scanned, this time for a trial \times consonant interaction or a trial \times hemisphere \times consonant interaction. The results revealed no significant trial \times consonant interactions; thus it appeared that bilateral /b/-/d/ discrimination was not affected by the use of synthetic syllables versus natural syllables. However, significant trial \times consonant \times hemisphere interactions were found for Factor 3 and Factor 8 [$F(1,11) = 4.84$, $p = .0500$; $F(1,11) = 6.77$, $p = .0401$; respectively]. Comparisons were made between /b/ and /d/ in the left hemisphere and the right hemisphere for either type of syllable using an a priori F statistic (Kirk, 1982).

For Factor 8, no significant differences were found for /b/ versus /d/ in *synthetic* syllables in the left hemisphere [$F(1,11) = 2.220$, $p = .162$] or the right hemisphere [$F(1,11) = 2.146$, $p = .169$]. Similarly, significant differences were not found between /b/ and /d/ in *natural* syllables in the left hemisphere [$F(1,11) = .314$, $p = .592$] or the right hemisphere [$F(1,11) = 1.296$, $p = .279$].

For Factor 3, however, a somewhat different pattern emerged. In this trial \times consonant \times hemisphere interaction, pictured in Figure 6, there were significant differences between /b/ and /d/ for both hemispheres and for both the synthetic- and natural-syllable trials [left-

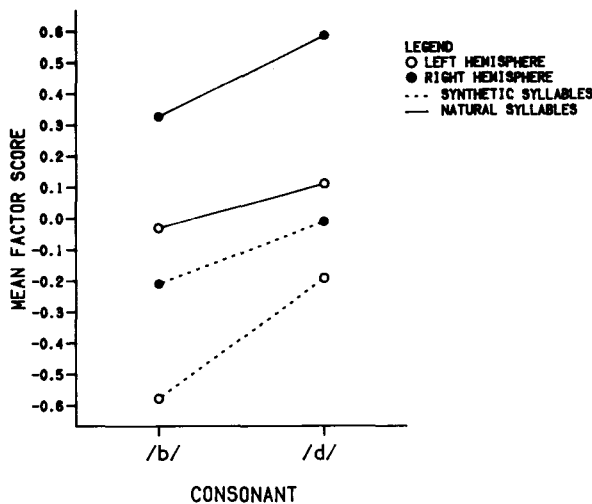


Figure 6. The trial \times consonant \times hemisphere interaction of Factor 3, Analysis 2. All comparisons between /b/ and /d/ were significant at the $\alpha < .05$ level.

hemisphere synthetic /b/ vs. /d/: $F(1,11) = 30.787$, $p = .0003$; right-hemisphere synthetic /b/ vs. /d/: $F(1,11) = 8.21$, $p = .0148$; left-hemisphere natural /b/ vs. /d/: $F(1,11) = 5.117$, $p = .0465$; right-hemisphere natural /b/ vs. /d/: $F(1,11) = 14.045$, $p = .0035$]. Although the degree of difference between /b/ and /d/ was larger in some cases than in others, the /b/-/d/ difference in each hemisphere did not appear to be a function of the two levels of the trial variable (synthetic syllables vs. natural syllables). Thus, it appeared that the pattern of hemispheric involvement in discriminating /b/ from /d/ was not significantly different, in either Factor 3 or Factor 8, for synthetic and natural syllables. These results suggest that synthetic syllables are valid alternatives to natural speech, at least for perceptual research involving voiced stop-bilabial and alveolar consonants. These results are also in agreement with behavioral data that show no significant differences between the accuracy of listeners' perceptual judgments of synthetic syllables versus natural syllables.

DISCUSSION

The results of Analysis 1 of this study were in agreement with those obtained by Molfese (1980) and Molfese and Schmidt (1983). That is, when CV syllables based on two stop consonants and three vowels were combined with nonspeech analogs, the consonants appeared to be discriminated bilaterally at a latency of 150–170 msec post stimulus onset and unilaterally in the left hemisphere at latencies of 215, 290, and 450–460 msec.

Although Analysis 1 of this investigation was not a direct replication of Molfese (1980) or Molfese and Schmidt (1983), the similarities in AER results were striking. The centroids from all three studies were remarkably alike in terms of waveshape (except for the final 150 msec) and latencies of major peaks. Furthermore, factor latencies associated with unilateral and bilateral discrimination were also similar among studies.

Importantly, these similarities occurred despite a variety of stimulus, methodological, and task differences between laboratories. In terms of stimulus differences, the present investigation used five-formant synthetic syllables, whereas the Molfese research used three-formant syllables; stimulus intensity was lower in the present experiment (62 dB vs. 80 dB); and rise time was longer (30 msec vs. 4 msec). In addition, this experiment included natural speech syllables as well as synthetic syllables; the nonspeech analogs were markedly different in bandwidth and duration; and, in the present research, the AER data were based only on trials in which the /b/ or /d/ consonant had been correctly identified. Methodological differences included use of different reference electrode sites (contralateral earlobes in this research, linked earlobes in the Molfese research) and different active electrode sites (T3 and T4 only in the present research; T3, T4, P3, P4, T5, and T6 in the later Molfese studies). Finally, there were a number of task variables that differed

between this and other research. In this experiment, subjects were asked to indicate which consonant they heard by lifting a finger on both hands, whereas in the Molfese research they were asked to identify the syllables silently (Molfese & Schmidt, 1983). Subjects in the present research were given instructions and training to identify the synthetic syllables and nonspeech analogs; Molfese's subjects were not. And subjects in the present experiment heard only one stimulus class at a time (synthetic syllables, natural syllables, chirps), whereas subjects in the Molfese research heard a random mixture of syllables and nonspeech analogs.

In contrast, in Analysis 2, the results were not in agreement with those obtained by Molfese (1980), Molfese and Schmidt (1983), or Analysis 1 of this experiment. When responses to the more difficult (i.e., nonspeech analog) stimuli were eliminated from the analysis, the previously obtained bilateral results remained unchanged, but significant unilateral discrimination was *not* observed. The difference in results between studies may be related to several variables. First, it may be that unless a task is somehow difficult or requires a certain level of attention, hemispheric asymmetry is not easily measurable. This hypothesis is supported by dichotic-listening research, in which making the perceptual task more difficult, such as reducing the duration of a stimulus or increasing its complexity, resulted in greater hemispheric asymmetry (Godfrey, 1974; Kasischke, 1979). The results of Eason et al. (1969), Harter and Salmon (1972), Poon et al. (1976), and Ritter et al. (1972) also support the idea that difficulty of the stimuli and/or required attention increase the measurability of AERs (in terms of increased amplitude of response and latency).

A second possible factor related to the inconsistently observed unilateral response may be the two-choice paradigm utilized in this research, in Molfese (1980), and in Molfese and Schmidt (1983). Although in all of these studies there were a total of six stimuli presented (2 consonants \times 3 vowels), subjects were required to distinguish only between one consonant and the other. It is possible therefore, that the similarities in results noted among the three studies, *and* the lack of a measurable unilateral response with syllable stimuli only, might have been due to experimental design.

This second hypothesis is supported by the results of Molfese (1984), in which the three consonants /b, d, g/ in one vowel context (speech-like stimuli only) were utilized. The results of this study, in contrast to earlier studies, showed that considerable discrimination took place in the right hemisphere. For example, one factor, with a primary latency of 295 msec, was associated with right-hemisphere differentiation between /b/-/g/ and between /d/-/g/. However, only the left hemisphere was found to differentiate between all possible pairs of consonants—and without the inclusion of the more difficult nonspeech analogs. In addition, although bilateral discrimination between /b/-/g/ and /d/-/g/ was associated with four out of six factors, no single bilateral process

was isolated that was sensitive to all possible consonant pairs (Molfese, 1984). Thus, it appears that experimental design, especially number of elements to be discriminated, is important in determining specific AER results. It is possible that a two-element distinction (with speech stimuli only) is not difficult enough to require unilateral processing, whereas a three-element choice may require just such involvement.

The results of this and other cited research emphasize the importance of taking experimental design and task variables into account before generalizing AER results to "speech perception." The latencies associated with various types of processing and even hemispheric involvement patterns appear to be dependent to a large extent on the particular task the subject is asked to do.

Finally, the use of synthetic speech stimuli appears to be an acceptable alternative to natural speech stimuli. No significant differences were found between the two types of syllables in consonant perception.

REFERENCES

- CHAPMAN, R., MCCRARY, J., BRAGDON, H., & CHAPMAN, J. (1979). Latent components of event-related potentials functionally related to information processing. In J. Desmedt (Ed.), *Progress in clinical neurophysiology: Vol. 6. Cognitive components in cerebral event-related potentials and selective attention*. Bergen, The Netherlands: Basel-Karger.
- CURRY, S. H., PETERS, J. F., & WEINBERG, H. (1978). Choice of active electrode site and recording montage as variables affecting CNV amplitude preceding speech. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 275-279). Washington, DC: U.S. Environmental Protection Agency.
- CUTTING, J. (1974). Two left hemisphere mechanisms in speech perception. *Perception & Psychophysics*, *16*, 564-570.
- DIXON, W. (1981). *BMDP statistical software*. Berkeley, CA: University of California Press.
- EASON, R., HARTE, M., & WHITE, C. (1969). Effects of attention and arousal on visually evoked cortical potentials and reaction time in man. *Physiology & Behavior*, *4*, 283-289.
- FRIEDMAN, D., SIMSON, R., RITTER, W., & RAPIN, I. (1975). Cortical evoked potentials elicited by real speech words and human sounds. *Journal of Electroencephalography & Clinical Neurophysiology*, *38*, 13-19.
- GODFREY, P. (1974). Perceptual difficulty and the right ear advantage for vowels. *Brain & Language*, *1*, 323-325.
- GOFF, G., MATSUMIYA, Y., ALLISON, T., & GOFF, W. (1977). The scalp topography of human somatosensory and auditory evoked potentials. *Journal of Electroencephalography & Clinical Neurophysiology*, *42*, 57-76.
- GOFF, W. (1974). Human average evoked potentials: Procedures for stimulating and recording. In R. Thompson & M. Patterson (Eds.), *Methods in physiological psychology: Vol. 1. Bioelectric recording techniques: Part B. Electroencephalography and human brain potentials*. New York: Academic Press.
- GOFF, W., MATSUMIYA, Y., ALLISON, T., & GOFF, G. (1969). Cross-modality comparisons of averaged evoked potentials. In E. Donchin & D. Lindsley (Eds.), *Average evoked potentials: Methods, results and evaluations* (NASA SP-191; pp. 95-141). Washington, DC: Government Printing Office.
- HARTE, M., & SALMON, L. (1972). Intra-modality selective attention and evoked cortical potentials to randomly presented patterns. *Journal of Electroencephalography & Clinical Neurophysiology*, *32*, 605-613.
- JASPER, H. (1958). The ten-twenty electrode system of the International Federation of Societies for Electroencephalography: Appendix to the

- report of the committee on methods of clinical examination in electroencephalography. *Journal of Electroencephalography & Clinical Neurophysiology*, **10**, 371-375.
- KAISER, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200.
- KASISCHKE, K. (1979). *An investigation of hemispheric dominance for simple auditory stimuli*. Unpublished doctoral dissertation, Washington University, St. Louis.
- KEWLEY-PORT, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, **72**, 379-389.
- KIMURA, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, **47**, 191-196.
- KIRK, R. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.
- KLATT, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.
- LIBERMAN, A., COOPER, R., SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LOW, M., WADA, J., & FOX, M. (1974). Electrographic localization of certain aspects of language production in the human brain. *Journal of Electroencephalography & Clinical Neurophysiology*, **36**, 562-563.
- LOW, M., WADA, J., & FOX, M. (1976). Electroencephalographic localization of conative aspects of language production in the human brain. In W. McCallum & J. Knott (Eds.), *The responsive brain* (pp. 165-168). Bristol, England: Wright.
- MOLFESE, D. (1978). Left and right hemispheric involvement in speech perception: Electrophysiological correlates. *Perception & Psychophysics*, **23**, 237-243.
- MOLFESE, D. (1980). The phoneme and the engram: Electrophysiological evidence for the acoustic invariant in stop consonants. *Brain & Language*, **9**, 372-376.
- MOLFESE, D. (1984). Left hemisphere sensitivity to consonant sounds not displayed by the right hemisphere: Electrophysiological correlates. *Brain & Language*, **22**, 109-127.
- MOLFESE, D., & SCHMIDT, A. (1983). An auditory evoked potential study of consonant perception in different vowel environments. *Brain & Language*, **18**, 57-70.
- MOWERY, G., & BENNETT, A. (1957). Some technical notes on monopolar and bipolar recording. *Journal of Electroencephalography & Clinical Neurophysiology*, **9**, 377.
- NEVILLE, H. (1980). Event-related potentials in neuropsychological studies of language. *Brain & Language*, **11**, 300-318.
- OLDFIELD, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, **9**, 97-113.
- PETERSON, G., & BARNEY, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.
- PICTON, T., HILLYARD, S., GALAMBOS, R., & SCHIFF, M. (1971). Human auditory attention: A central or peripheral process? *Science*, **173**, 531-533.
- POON, L., THOMPSON, L., & MARSH, G. (1976). Average evoked potential changes as a function of processing complexity. *Psychophysiology*, **13**, 43-49.
- RITTER, W., SIMSON, R., & VAUGHN, H. (1972). Association cortex potentials and reaction time in auditory discrimination. *Journal of Electroencephalography & Clinical Neurophysiology*, **33**, 547-555.
- SCHWAB, E. (1981). *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral dissertation, State University of New York, Buffalo.
- SCHWENT, V., & HILLYARD, S. (1975). Evoked potential correlates of selective attention with multi-channel auditory inputs. *Journal of Electroencephalography & Clinical Neurophysiology*, **38**, 131-138.
- SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59-63.
- WEISS, H., & HOUSE, A. (1973). Perception of dichotically presented vowels. *Journal of the Acoustical Society of America*, **53**, 51-58.
- WOLPAW, J., & WOOD, C. (1982). Scalp distribution of human auditory evoked potentials: I. Evaluation of reference electrode sites. *Journal of Electroencephalography & Clinical Neurophysiology*, **54**, 15-24.

NOTES

1. The choice of reference electrode site is a controversial issue in electrophysiological literature. If one is attempting to use a monopolar recording paradigm, the reference electrode site should ideally have the same myogenic, artifactual, and interference potentials as the active electrode site, but none of the evoked activity. If this were the case, the unwanted "noise" potentials from the two sites would be cancelled through the use of a differential amplifier, leaving only the evoked activity to be recorded. Unfortunately, such an ideal reference site is not universally agreed upon. Many investigators use linked earlobes, although according to Mowery and Bennett (1957), the ear with the lowest resistance may be the actual referent. G. Goff et al. (1977), W. Goff (1974), and W. Goff et al. (1969) recommend the use of the earlobe contralateral to the stimulus as "the best compromise as a common reference point to compare AEPs across modalities" (W. Goff, 1974, p. 118), and found that this "indifferent" location was "freer from potentials evoked by the three kinds of stimulation than the right [ipsilateral] ear, the bridge or tip of the nose, mastoid processes, or chin" (G. Goff et al., 1977, p. 59). Nevertheless, a variety of reference sites continue to be used, including linked earlobes (Molfese, 1980; Molfese & Schmidt, 1983), contralateral earlobes (Low, Wada, & Fox, 1974, 1976), a single (right) earlobe (Schwent & Hillyard, 1975), ipsilateral mastoids (Picton, Hillyard, Galambos, & Schiff, 1971) a single (right) mastoid (Neville, 1980, Experiment 1), or the tip of the nose (Friedman, Simson, Ritter, & Rapin, 1975). In fact, a recent article by Wolpaw and Wood (1982) suggested that *all* reference sites on the head or face are contaminated by evoked activity in most subjects (and thus not truly inactive), and that a sternovertebral reference appears to be the best choice for AER recording.

Despite this controversy, studies in which reference electrode sites have been manipulated suggest that specific evoked responses may be robust to such changes. For example, G. Goff et al. (1977) repeated an AER topographic study, first using the right ear (contralateral) as the reference site, and then (for 3 subjects) using a noncephalic indifferent. They found no difference in topographic results. In a study of hemispheric asymmetry preceding speech, Curry, Peters and Weinberg (1978) examined contingent negative variation (CNV) amplitude when linked earlobes versus contralateral earlobes were used as the referents. These investigators found no differences in CNV amplitude or asymmetry results, regardless of reference-electrode montage. Visual inspection of Curry et al.'s waveforms in the two conditions revealed no marked differences in configuration. Thus, the use of contralateral earlobes for reference sites in this or any other research appears to be an acceptable alternative to linked earlobes. Furthermore, there is no compelling reason to believe that the results of such studies would not be comparable.

2. The study described in this paper was part of a larger research design, which included comparisons between chirps with frequency instructions and chirps with speech instructions. Such comparisons are not within the scope of this paper; however, in order to report the statistical procedures accurately, these chirps must be noted as the fourth trial condition. Results involving frequency chirps will not be reported in this manuscript. In the remainder of the paper, the term *chirps* will refer exclusively to *chirps with speech instructions*.