BRAIN AND LANGUAGE 9, 372–376 (1980)

# NOTES AND DISCUSSION

## The Phoneme and the Engram: Electrophysiological Evidence for the Acoustic Invariant in Stop Consonants

DENNIS L. MOLFESE

*Southern Illinois University at Carbondale*

Investigators in nearly three decades of research have been unable to identify the acoustic invariance in stop consonants—the cue that enables us to identify a particular consonant independent of context. Although the second formant transition is known to be a major cue for most consonants (Delattre, Liberman, & Cooper, 1955), the frequencies of these transitions change as a function of the vowels which precede or follow these transitions. Nevertheless, consonant identification remains unchanged. The initial consonants of the syllables /di/ and /du/ are perceived as the same phoneme, /d/, even though the acoustic cue for this consonant is different for the two syllables. The second formant transition cue rises from approximately 2200 to 2600 Hz for /di/ while it falls from 1200 to 700 Hz for /du/. For consonants, then, "what is perceived as the same phoneme is cued, in different contexts, by features that are vastly different in acoustic terms" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

The present paper reports for the first time the identification of neuroelectrical responses which reflect the processing of phonemes within the cortex. Auditory evoked responses (AERs), those portions of the EEG activity which can be time-locked to the occurrence of specific stimulus events, were recorded from temporal and parietal sites over the left and right hemispheres in response to a series of consonant–vowel syllables in which the consonants and vowels were varied. These measures have been found to reflect stimulus and subject variables in numerous studies (Molfese, 1978; Regan, 1972).

This research was supported in part by a grant (6-21874) from The National Foundation/March of Dimes. The author wishes to thank Albert Schmidt and Roland Erwin for their assistance during the subject testing and analysis phases of this work. Requests for reprints should be addressed to Dennis L. Molfese, Department of Psychology, Southern Illinois University at Carbondale, Carbondale, IL 62901.



0093-934X/80/020372-05$02.00/0

## METHOD

*Subjects.* Twenty right-handed Caucasian females, ranging in age from 18 to 24 years, participated in the present study.

*Stimuli.* Twelve three-formant consonant–vowel syllables synthesized by Dr. James Cutting at Yale University and the Haskins Laboratories were presented as stimuli. All syllables consisted of an initial rapid frequency transition 50 msec in duration, followed by 300 msec steady state formants. The 12 stimuli represented two different stimulus sets. The first set consisted of six $CV_{NF}$ syllables which corresponded to the consonant–vowel syllables /bi, bae, bɔ, gi, gae, gɔ/. The $CV_{NF}$ stimuli were characterized by normal formant structure with bandwidth of 60, 90, and 120 Hz for formants 1, 2, and 3, respectively. The second set, $CV_{SF}$, differed from the first in formant bandwidth. All formants consisted of 1-Hz sine waves whose frequencies matched the middle values of the $CV_{NF}$ formants. For additional information concerning the nature of these stimuli, the reader is referred to Cutting (1974).

Sixteen different orderings of the 12 stimuli were recorded on one channel of a stereo tape recorder (Sony Model TC-353). The time interval between stimuli was randomly varied from 4 to 8 sec in order to reduce the effects of expectation and habituation. A pulse, time-locked to the onset of each stimulus, and recorded on a second channel of the tape, served to identify the onset of each stimulus and the subsequent auditory evoked response for a PDP-12 computer.

*Procedure.* Subjects were tested individually while seated in a reclining chair in a sound-dampened and electrically shielded chamber. The stimuli were presented through a speaker centered approximately 1 m directly above the individual's head. The stimulus peak intensity at the subject's ears was 80 db (A). Grass silver electrodes were placed on the scalp over the temporal and parietal regions of the left hemisphere at $T_3$, $T_5$, and $P_3$ and over corresponding areas of the right hemisphere at $T_4$, $T_6$, and $P_4$ (Jasper, 1958). Each scalp electrode was referred to linked ear lobes. Electrode impedances for each side of the head were checked and recorded before and after the testing session for each subject. The mean values of these resistances were below 5 kohm before and at the end of the 15-min testing session. Resistances between leads were maintained within 1.0 kohm. The electrodes were connected to Analogue Devices isolation amplifiers (Model 284J) which in turn were connected to modified Tektronic AM502 differential amplifiers with the bandpass flat between 0.1 and 30 Hz and with gain settings at 20 K. Auditory evoked responses (AER) elicited in response to the stimuli were recorded on a Vetter modified cassette FM tape recorder (Model C-8) for later analyses.

## RESULTS

The AERs from each subject served as the dependent measure in the present study. These were digitized and averaged on a PDP-12 computer using a modified version of "averager" (Decus No. 12-84). Averages were based on 16 repetitions of each stimulus.

Analyses were performed on the AERs using principal component analysis and analysis of variance techniques described by Molfese (1978). The reader is referred to this paper for a more detailed description of the analyses used in the present study. The data set consisted of the digitized amplitude values of 1440 averaged evoked responses which were obtained from three sites over each hemisphere of 20 adults for the 12 different stimuli. An input data matrix was obtained from the 1440 averaged AERs by taking the amplitude values at each of the 100 time points at intervals of

5 msec over the 500-msec period following stimulus onset. These data were then submitted to a principal components analysis (BMDP4M) using the BMDP77 program package (Brown, 1977). This program first transformed the data matrix into a correlation matrix. The principal components analysis was then applied to a $100 \times 100$ matrix which consisted of the product-moment correlations computed for each pair of time points. Factors which met the Cattell Scree Test criterion were retained (Cattell, 1966). In addition, only factors which accounted for at least as much variance in the data as any one of the original 100 variables were retained. Seven factors which accounted for 70% of the total variance were isolated. These were then rotated using the normalized varimax criterion which preserved the orthogonality among the factors while improving their distinctiveness. The factors each consisted of 100 factor loadings which corresponded to the 100 time points. The factor loadings reflected the association of the factors to the original variables (time points). Factor scores (gain factors) were then computed for each of the 1440 original AERs for each of the seven rotated principal components.

Independent analyses of variance for Consonants (2) × Vowels (3) × Formant Structure (2) × Hemispheres (2) × Sites (3) were performed on the factor scores for each factor in order to determine if any of the factors varied systematically as a function of stimulus manipulations or recording sites. A $p < .01$ criterion level of significance was set. Two factors were found in this manner to reflect changes in the AER components which were related to phoneme differences independent of vowel environments. Only these factors are presented in the present report. The centroid (the averaged auditory evoked brain potential for the entire data set) and these two factors are presented in Fig. 1. The centroid is characterized by a small positive wave which peaked 30 msec after stimulus onset ($P_{30}$). This component was followed by a major negative peak at 110 msec ($N_{110}$), a large positive wave ($P_{200}$), and several smaller negative peaks ($N_{270}$ and $N_{450}$). Factor 1, which accounted for 11.10% of the total variance, was distinguished by waves with peak latencies at 10, 215, 290, and 460 msec following stimulus onset. Factor 3 (10.01% of total variance) contained a single peak that occurred 170 msec following stimulus onset. A Hemispheres × Consonants interaction characterized Factor 1, $F(1, 19) = 19.12$, $p < .001$. A Scheffé test of this interaction indicated that only the left hemisphere differentiated between consonants on the basis of phoneme class. A significant main effect for consonants, $F(1, 19) = 9.77, p < .01$, characterized Factor 3. The group averaged AERs elicited by the /b/ and /g/ syllables are presented in Fig. 2. The major differences between the two waveforms depicted in this figure reflect the contribution of Factor 1 to the first 50 msec and the last 300 msec of the AERs and the contribution of Factor 3 to the AERs between 100 and 200 msec.

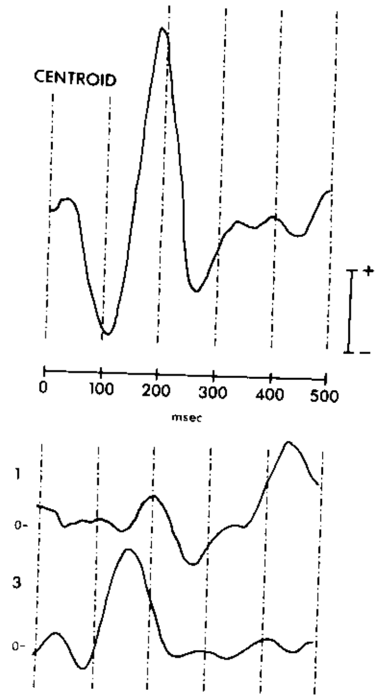


FIG. 1. The centroid or group average AER for the entire data set and the two factors obtained by the varimaxed principal components analysis. Stimulus onset was at 0. AER duration is 500 msec. The calibration marker is 2 $\mu$V with positive up.

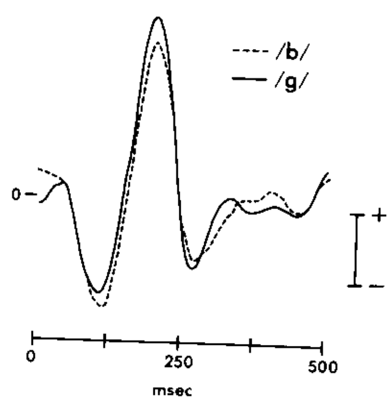


FIG. 2. The group average AERs elicited for the /b/ (dashed line) and /g/ (solid line) syllables. These averages are collapsed across all vowel conditions. The calibration marker is 2 $\mu$V with positive up.

## DISCUSSION

The findings reported here represent the first evidence that the consonant phoneme is processed as a distinct and independent unit in the cortex. Several regions of the cortex were found to differentiate between consonants independent of the context surrounding these phonemes. Although the acoustic frequencies for the consonant transitions changed when the consonant was combined with different vowels, two components of the brain's electrocortical response were found to distinguish between phoneme classes rather than between acoustic differences. It would appear that at some level within the cortex of man the phoneme may in fact be the basic perceptual unit of language.

## REFERENCES

Brown, M. B. (Ed) 1977. *Biomedical computer programs—P series*. Berkeley, CA: Univ. of California Press.

Cattell, R. B. 1966. The Scree test for the number of factors. *Multivariate Behavioral Research*, **1**, 245.

Cutting, J. E. 1974. Two left hemisphere mechanisms in speech perception. *Perception and Psychophysics*, **16**, 601–612.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. 1955. Acoustic loci and transitional cues for consonants. *Journal of Acoustical Society of America*, **27**, 769–773.

Jasper, H. 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. *The Journal of Electroencephalography and Clinical Neurophysiology*, **10**, 371.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review*, **74**, 431–461.

Molfese, D. L. 1978. Left and right hemisphere involvement in speech perception: Electrophysiological correlates. *Perception and Psychophysics*, **23**, 237–243.

Regan, D. 1972. *Evoked potentials in psychology, sensory physiology, and clinical medicine*. New York: Wiley.