Auditory Evoked Potentials During Speech Perception

Charles C. Wood, William R. Goff, and Ruth S. Day

ABSTRACT

Neural responses evoked by the same binaural speech signal were recorded from ten right-handed subjects during two auditory identification tasks. One task required analysis of acoustic parameters important for making a linguistic distinction, while the other task required analysis of an acoustic parameter which provides no linguistic information at the phoneme level. In the time interval between stimulus onset and the subjects' identification responses, evoked potentials from the two tasks were significantly different over the left hemisphere but identical over the right hemisphere. These results indicate that different neural events occur in the left hemisphere during analysis of linguistic versus nonlinguistic parameters of the same acoustic signal.

The relation between an acoustic speech signal and its phonetic message appears to be a complex and highly efficient code, which requires a specialized linguistic "decoder" for its perception (Liberman et al., 1967; Mattingly and Liberman, 1969; Studdert-Kennedy et al., 1970; Liberman, 1970). Dichotic listening experiments using normal (Kimura, 1961b, 1964, 1967; Shankweiler and Studdert-Kennedy, 1967; Curry, 1967; Curry and Rutherford, 1967; Kimura and Folb, 1968, Darwin, 1969a,b; Day and Cutting, 1970a,b; Studdert-Kennedy and Shankweiler, 1970) and brain-damaged subjects (Kimura, 1961a; Shankweiler 1966; Sparks and Geschwind, 1968; Milner et al., 1968; Schulhoff and Goodglass, 1969; Sparks et al., 1970) have further suggested that the specialized neural mechanisms required for the perception of speech are lateralized in one cerebral

---

+ Neuropsychology Laboratory, Veterans Administration Hospital, West Haven, and Department of Psychology, Yale University, New Haven.
++ Neuropsychology Laboratory, Veterans Administration Hospital, West Haven, and Departments of Neurology and Psychology, Yale University, New Haven.
+++ Haskins Laboratories, New Haven, and Department of Psychology, Yale University, New Haven.

Acknowledgments: Supported by the Veterans Administration, U.S. Public Health Service NTMH Grant M-05286, and NSF Grants GB-3919 and GB-5782 to W.R. Goff, and by NICHD Grant HD-01994 to the Haskins Laboratories. We thank T. Allison, F.S. Cooper, M.W. Donald, D. Shankweiler, and M. Studdert-Kennedy for comments on the manuscript, and J.E. Cutting, T.C. Fisher, and J. Jasiorkowski for technical assistance.
hemisphere, usually the left. This interpretation is consistent with
clinical analyses of language disorders following brain damage and may
be related to anatomical differences between left and right temporal lobes
(Geschwind and Levitsky, 1968). In a recent review of hemispheric speciali-
ization for speech perception, Studdert-Kennedy and Shankweiler (1970, p. 579)
concluded that "specialization of the dominant hemisphere in speech percep-
tion is due to its possession of a linguistic device.... While the general
auditory system common to both hemispheres is equipped to extract the audi-
tory parameters of a speech signal, the dominant hemisphere may be special-
ized for the extraction of linguistic features from those parameters."

Despite the large body of behavioral and clinical evidence for speciali-
ization of one hemisphere in speech perception, there is no evidence which
clearly distinguishes neural activity specifically related to linguistic
processing from that which occurs during the processing of any auditory
stimulus. Empirical evidence for such a distinction requires a direct

---
1 For a recent review see Geschwind (1970).
2 Three experiments concerning neural activity evoked by speech sounds have
been reported. Greenberg and Graham (1970) reported larger amplitudes of
the evoked potential's "largest amplitude spectral component" from left-
than right-hemisphere locations during a CV syllable learning task. No
statistical evidence was included to show that the obtained results differed
significantly from those expected by chance variation. Roth et al. (1970)
reported no significant differences in activity recorded at the vertex to
"sense and non-sense" monosyllables. In a paper published after the pre-

current experiment was submitted, Cohn (1971) reported "differential cerebral
processing of noise and verbal stimuli." Cohn's major result was a "prom-
inent positive-going peak with a latency of around 14 msec in the right
brain derivation" in response to "click" stimuli generated by 10-msec pulses
but not in response to "single syllable words" generated and presented in
an unspecified manner. There are three major difficulties with the Cohn
experiment: 1) No statistical evidence was presented to demonstrate that
the obtained results differed significantly from chance variation. 2) Cohn's
"verbal" and "noise" stimuli differed in many acoustic parameters such as
duration, frequency composition, rise-time, total amplitude, and amplitude
contour. Differences in neural activity evoked by such stimuli could be
related to any or all of such acoustic differences, none of which need
have any direct bearing upon the issue of speech versus nonspeech perception
which Cohn wished to address. 3) It is possible that auditory evoked po-
etentials of 14-msec latency are of nonneural origin. In previous reports
of potentials recorded under conditions similar to those used by Cohn, the
shortest latency potentials recordable from the human scalp and considered
to be of neural origin do not occur until approximately 30 msec (Mast, 1965;
Ruhm et al., 1967; Goff et al., 1969). Potentials in the 14-msec latency
range have been considered to be nonneural artifacts (Bickford et al., 1964;
Mast, 1965; Goff et al., 1969).

An experiment by MacAdam and Whitaker (1971) dealt with the question of
hemispheric specialization of speech production. They reported slow poten-
tials, distributed largest over the left hemisphere, occurring up to 1 sec
before the production of polysyllabic words. Symmetrically distributed po-
etentials were reported before the production of similar nonspeech gestures.
comparison of neural activity during linguistic and nonlinguistic process-
ing conditions with other sources of variation in neural activity elimin-
ated between conditions. We have therefore compared neural activity evoked
by the same consonant-vowel syllable during two auditory identification tasks:
one that required analysis of acoustic parameters which provide linguistic
information (Stop Consonant Task) and one that required analysis of acoustic
parameters which provide no linguistic information at the phoneme level
(Fundamental Frequency Task). For convenience, we shall use the terms "lin-
guistic and nonlinguistic parameters" to refer to those acoustic parameters
that do and do not, respectively, provide linguistic information at the pho-
neme level.

Stop Consonant Task. Subjects were required to indicate which of two
possible stimuli had occurred on each trial: /ba/ or /da/. The stimuli
were generated by the Haskins Laboratories parallel resonance synthesizer
and were prepared to be identical in duration (300 msec), initial fundamental
frequency (F0 = 104 Hz), frequency contour (falling), and intensity contour
(falling). Thus, the two syllables differed only in those acoustic cues
important for distinguishing between voiced stop consonants, namely the di-
rection and extent of the second (Liberman et al., 1954; Delattre et al.,
1955) and third (Harris et al., 1958) formant transitions. Stop consonants
were selected for the linguistic task since they appear to be the most highly
encoded of all phonemes (Liberman et al., 1967; Mattingly and Liberman, 1969;

Fundamental Frequency Task. Again subjects were required to indicate
which of two possible stimuli had occurred on each trial. In this task, how-
ever, the two stimuli had identical linguistic information, namely formant
transitions appropriate for the syllable /ba/. They differed only in funda-
mental frequency: /ba/-low (initial F0 = 104 Hz) versus /ba/-high (initial
F0 = 140 Hz). Both stimuli were 300 msec in duration and had frequency and
intensity contours matched to those of stimuli in the Stop Consonant Task.
Variations in fundamental frequency were selected for the nonlinguistic task
since absolute fundamental frequency provides little or no linguistic infor-
mation at the phoneme level in English. Thus, the two tasks employed three
acoustic stimuli, with the syllable /ba/-low (initial F0 = 104 Hz) common to
both tasks and used for comparison of evoked potentials. Spectrograms of
the three stimuli are shown in Figure 1 arranged according to identification
task.

Ten right-handed subjects (ages 18-20) were each tested during two sep-
arate sessions.3 Both sessions consisted of six blocks of sixty-four stimuli,
three blocks each of the Stop Consonant and Fundamental Frequency Tasks. A
block of sixty-four stimuli contained thirty-two each of the two possible
stimuli for that task, presented in random order at 5-sec interstimulus
intervals. The two tasks were presented in alternating order during each
session. Five subjects began session 1 with the Stop Consonant Task and
session 2 with the Fundamental Frequency Task; the remaining five subjects

3Right-handed subjects were selected for this experiment since most are left-
hecisphere dominant for language. See, for example, Zangwill (1960), Branch
et al. (1964), Milner (1967), and Rossi and Rosadini (1967).
Spectrograms of the Three Stimuli

Note: The syllable /ba/-low ($F_0 = 104$ Hz) was identical for both tasks. In the Stop Consonant Task, stimuli differed only in those acoustic cues important for distinguishing between voiced stop consonants (direction and extent of the second and third formant transitions). In the Fundamental Frequency Task, stimuli had identical linguistic information and differed only in fundamental frequency.
began the two sessions in the reverse order. Subjects were required to indicate which of the two possible stimuli they heard on each trial as soon as possible following stimulus onset. In both tasks, subjects pressed button 1 with the right index finger when they heard /ba/-low and button 2 with the right middle finger when they heard the other stimulus. Thus, both identification tasks contained an identical acoustic stimulus (/ba/-low), which occurred an equal number of times (thirty-two per run of sixty-four), with equal presentation probability on each trial (p = .50), and which required an identical motor response (pressing button 1 with the right index finger). Before session 1, subjects were asked to listen to the three acoustic stimuli and report what they heard. All subjects correctly identified each of the three syllables. They were then allowed to practice each task under conditions identical to those of the experiment until reaction times were stable. All subjects made fewer than five errors per run of sixty-four stimuli, and errors did not differ significantly between tasks. Therefore error scores will not be considered.

Electrical activity was recorded from temporal and central 10-20 system (Jasper, 1958) scalp locations over the left hemisphere (T3 and C3) and from corresponding locations over the right hemisphere (T4 and C4), each referred to a linked-ear reference using silver disc electrodes. Impedances of all electrodes were monitored regularly during each session and were less than 2.5 kilohms paired with the linked-ear reference. Particular care was taken to equalize impedances of the two ear reference electrodes: in all subjects both reference electrodes were equal at less than 3.0 kilohms, paired with each of the other electrodes.

Subjects were seated comfortably in a sound-attenuating and electrically shielded chamber illuminated at moderate intensity. EEG was recorded with a Grass Model 7 polygraph using Grass Model 7P5A wide-band A.C. EEG pre-amplifiers (system gain = $2 \times 10^4$) and was monitored visually throughout each run. Half-amplitude low- and high-frequency settings were 0.3 Hz and 500 Hz, respectively. Amplified signals were entered into a LINC computer for analog-to-digital conversion and signal averaging. Sampling epochs were 490 msec with 256 time points per epoch. The LINC controlled the stimulus presentation order, averaged evoked potentials separately for each of the two stimuli in each task, and stored the averaged responses on magnetic tape for off-line data analysis. Subjects' identification responses and reaction times were recorded using a Beckman-Berkeley Model 7531R Universal Counter-Timer.

The synthetic stimuli were played to the subjects from a Precision Instrument FM tape recorder (Frequency response: $\pm$ 0.5 db, DC to 10 KHz at 30 ips). They were presented binaurally at 65 db SL against a 30-db white noise through a Grason-Stadler Model 829D electronic switch to G. C. Electronics earplug-type earphones. The timing of all events, including the initiation of LINC sampling epochs, was controlled by pulses on a separate channel of the FM tape recorder synchronized with stimulus onset.

---

4The 256 time points were distributed throughout the 490-msec epoch at three sampling rates: 1 point every 0.5 msec for the first 60 points, 1 point every 1 msec for the next 66 points, and 1 point every 3 msec for the remaining 130 points.
Averaged potentials evoked by the identical stimulus in both tasks (/ba/-low) were combined across subjects to obtain averages of 1,920 responses for each task and electrode location as shown in Figure 2. Evoked potentials from the Stop Consonant and Fundamental Frequency Tasks are superimposed at each electrode location to facilitate visual comparison. Reaction times did not differ significantly between tasks according to a Wilcoxon test (Siegel, 1956) (Median ± Semi-interquartile Range: Stop Consonant = 502 ± 75 msec, Fundamental Frequency = 493 ± 70 msec; T = 15, N = 10, p > .10). To determine the statistical reliability of differences between evoked potentials from the two tasks, Wilcoxon tests (Siegel, 1956) were computed between evoked potentials at each of the 256 individual time points in the sampling epoch. Results of the statistical analyses are shown in Figure 2 below the evoked potentials at each of the four electrode locations. Upward deflections from baseline in the statistical traces indicate that the difference between evoked potentials at the time point was significant at the .01 level. For significance at the .01 level, the computation procedure for the Wilcoxon tests requires that the differences between evoked responses for a given time point occur in at least eight of the ten subjects.

In order to analyze evoked potentials during the identification processes required by the two tasks, the 490-msec evoked potential sampling epoch was empirically divided into the pre-response and motor response intervals shown in Figure 2. Since the identification process must be complete at or before the identification response is made, only the pre-response interval is appropriate for the analysis of evoked potentials during the identification process. Differences between evoked potentials during the motor response interval will be considered below.

---

5 This procedure was designed to determine a) the statistical reliability of differences between evoked potentials from the two tasks and b) the precise distribution of significant differences in time relative to stimulus onset and subjects' identification responses. Our procedure computed statistical significance for each of the 256 evoked response time sample points, using a standard non-parametric paired comparison technique (Wilcoxon matched-pairs signed-ranks test). At every sample point, the difference between amplitudes of responses from Stop Consonant and Fundamental Frequency Tasks was obtained for each of the ten subjects. The differences between tasks were then ranked and the Wilcoxon T statistic was calculated in the usual manner (Siegel, 1956). Thus, a value of the T statistic for the difference between evoked potentials from the two tasks was obtained for each of the 256 individual time points in each pair of responses.

6 On a single trial the motor identification response unambiguously ends the time interval during which the identification process must have occurred. However, in the average of large numbers of trials required for comparison of evoked potentials, the proper end of the "processing interval" is less clear. Our criterion for distinguishing the pre-response and motor response intervals was the time point after which 99 percent of the motor responses occurred. The 99 percent point was selected instead of the 100 percent point because it disregards those few trials with extremely short RTs which cannot be meaningfully related to the identification tasks.
Electrical Activity Evoked by the Identical Stimulus in Both Tasks /ba/-low (F₀ = 104 Hz)

![Graph showing electrical activity in the left and right hemispheres.](image)

*Note:* Averages of 1,920 responses for the Fundamental Frequency (solid lines) and Stop Consonant (dotted lines) Tasks are shown for left (T3 and C3) and right (T4 and C4) hemisphere locations. Positivity at the scalp electrode relative to the linked ear reference is upwards. Time scale: 500 msec. Vertical calibration: 5 uV. Results of the Wilcoxon statistical tests (Siegel, 1956) for each of the 256 individual sample points are shown below the evoked potentials at each electrode location. Upward deflections of the statistical traces indicate that the difference between evoked potentials at that individual time point is significant at the .01 level.
If the analysis of linguistic and nonlinguistic parameters of an acoustic signal consists of the same neural events, then evoked potentials should be the same (within the limits of normal variation) for both tasks during the pre-response interval. Evoked potentials from the right hemisphere (T4 and C4) were indeed identical for both tasks during pre-response interval, as shown in Figure 2. However, statistically significant differences in evoked potentials occurred at left-hemisphere locations (T3 and C3) during the same time interval. By chance variation, 1.77 significant time points would be expected at each location during the pre-response interval. At temporal and central locations over the left hemisphere 30 and 34 significant points were obtained, while 1 and 0 significant points were obtained at corresponding right-hemisphere locations. These results indicate that neural events in the right hemisphere were identical for both tasks during the pre-response interval, regardless of the task requirements. In contrast, different neural events occurred in the left hemisphere during the same time interval, depending upon whether the task required analysis of linguistic or nonlinguistic parameters of the acoustic signal.

We have been careful to eliminate factors which could produce artifactual differences in evoked potentials between tasks. There is, however, one additional source of possible artifact. Since the occurrence of a motor response (Karlin et al., 1970) and the speed of the that response (Bostock and Jarvis, 1970) can alter the neural activity evoked by sensory stimulation, it is possible that even nonsignificant differences in RT produced the results shown in Figure 2. To examine this possibility, the evoked potentials at each electrode location were recategorized. Instead of averaging the six Stop Consonant and six Fundamental Frequency blocks for each subject, the six fastest and six slowest RT blocks were averaged to maximize RT differences. Evoked potentials from the fast and slow RT blocks were then analyzed statistically in the same way as those in Figure 2.

If the evoked potential differences during the pre-response interval in Figure 2 were produced by nonsignificant differences in RT, then similar or larger differences should be produced by averaging the blocks with slowest and fastest RTs. Such a result did not occur. No more significant differences than would be expected by chance occurred at any electrode location during the pre-response interval: 1 significant point was obtained at each left-hemisphere location, and 1 and 2 significant points, respectively, were obtained at right-hemisphere locations. During the motor response interval, evoked potentials from the slow and fast RT blocks were significantly different in the same direction as those during the motor response interval in Figure 2. Thus, we cannot rule out the possibility that slight differences in RT may have produced the effects during the motor response interval shown in Figure 2. However, differences in RT could not have produced the significant differences in evoked potentials during the pre-response interval.

In summary, this experiment demonstrates that: 1) differences in neural responses evoked by the same speech signal occurred between tasks which required analysis of linguistic versus nonlinguistic parameters of that signal; 2) such differences occurred only at left-hemisphere locations; and 3) these differences are not related to differences in the acoustic signal, its presentation probability, the subjects' motor response, or reaction time. These results indicate that different neural events occur in the left hemisphere during analysis of linguistic versus nonlinguistic parameters of the same
acoustic signal. Further, they provide strong support for the idea that a unilateral neural mechanism is specialized to perform those linguistic processes necessary for speech perception.

REFERENCES


