Human Brain Specialization for Phonetic Attention

Szymanski, Michael, D.*, Yund, E. William, and Woods, David.L.

Department of Neurology (127),
University of California, Davis,
VA Northern California System of Clinics,
150 Muir Rd.
Martinez, CA 94553 USA

*Present address
of corresponding author:

Department of Radiology
Biomagnetic Imaging Laboratory
UC San Francisco
513 Parnassus Avenue S-362
San Francisco, CA 94143-0628
mszyman@mail.radiology.ucsf.edu
Abstract

The effects of auditory selective attention on event related potentials (ERPs) to speech sounds were examined in subjects attending to vowel-consonant-vowels (VCVs) in one ear while ignoring VCVs in the opposite ear. In one condition, subjects discriminated phonetic changes in the VC, CV, or both formant-transition regions. In another condition, they discriminated equally difficult intensity changes in the same VCV regions. Attention-related negative difference waves showed enhanced early and late components ($N_{de}$ and $N_{dl}$) during phoneme-discrimination conditions. Hemispheric asymmetries developed only during the $N_{dl}$ and were more pronounced during phoneme discrimination. The results suggest that auditory areas of both hemispheres are specialized for phonetic analysis, with hemispherically specialized mechanisms engaged primarily during the final stages of phoneme processing.

Key words: cortex, attention, speech, phoneme, intensity, $N_d$, negative difference wave, auditory, event-related potentials, ERPs.
Introduction

Human subjects show a remarkable ability to selectively attend to one voice while ignoring other, competing voices. The physiological basis of this “cocktail party phenomenon” remains poorly understood, but is thought to involve both the enhanced processing of attended stimuli and the gating of non-attended signals [1]. The physiological mechanisms of auditory attentional selection in humans have been extensively studied with event-related potentials (ERPs), where attention effects can be isolated as negative difference (Nd) waves by subtracting ERPs to unattended stimuli from the ERPs to the same stimuli when attended. The Nd has early (Nd_e) (50-200 ms) and late (Nd_l) (200-800 ms) sub-components [2].

Nd waves reflect the activation of neuronal populations in auditory cortex [3,4], and are influenced by the stimulus features to which the subject attends. For example, Nd waves associated with pitch and location processing have different scalp distributions [5,6] reflecting the modulation of distinct auditory cortical fields. Such feature-specific effects are likely to extend to speech processing. There is considerable evidence that extensive regions of human auditory cortex are specialized for the analysis of speech sounds [7] particularly in the left hemisphere [8]. This line of reasoning suggests that Nd waves might be particularly prominent during the linguistic processing of speech stimuli, and Nd asymmetries might be enhanced during conditions that emphasize linguistic analysis.

Previous ERP studies of attention to speech have provided a conflicting picture. Although initial experiments suggested that hemispheric asymmetries were evident during phonetic but not acoustic analyses of speech sounds [9], these results proved difficult to replicate [10]. Subsequent experiments reported Nd waves with similar amplitudes and scalp distributions in conditions where subjects attended to CV or tone sequences [11,12]. However, most of these experiments used simple and repetitive phonetic sequences, so that subjects could have successfully performed the phonetic task using acoustic strategies. Furthermore, the distinctive features of the CVs that required phonetic discrimination occurred at stimulus onset, where automatic orienting might be expected to direct some attention even to nominally unattended stimuli. In the current experiment, we analyzed Nd waves to phonetically complex vowel-consonant-vowel (VCV) stimuli in conditions where subjects detected deviations in intensity or in phonetic structure. In addition to the greater phonetic complexity of the VCV versus CV stimuli,
the consonant feature of the VCV occurs in the center of the stimulus rather than at stimulus onset. This avoids any confounding effects of automatic orienting by giving the listener time to (re)focus attention before the phonetically relevant portions of the signal occur.

**Materials and Methods**

Thirteen subjects (five males and eight females, ages 20-39 yrs) with normal hearing (thresholds 20 dB HL or less from 250 to 4000 Hz) each completed two recording sessions. They were paid for their participation and gave informed consent, according to DVA guidelines.

The VCV stimuli were synthesized at 20 kHz sampling rate with a Klatt speech synthesizer, KLSYN88a (MIT and Sensimetrics Corporation) in cascade mode with 1 ms update interval. Stimuli were delivered under the control of a PC via Etymotic ER-2 earphones. The intensity of the steady-state portion of the vowels in the VCV stimuli was 75 dB SPL.

The standard left ear stimulus was a female-voice /ibi/ that had a fundamental frequency (F0) of 220 Hz and vowel formant frequencies of 330 (F1), 2000 (F2), 3000 (F3) and 3600 Hz (F4). The standard right-ear male-voice /ibi/ differed only in F0 (156 Hz) and F1 (270 Hz). There were two types of consonant deviants: (1) phoneme deviants, where /b/ was changed to /g/, and (2) intensity deviants, where the intensity of /b/ was increased (referred to as /B/). The end points for the formant transitions for /b/ were 1600 (F2), 2300 (F3) and 2800 (F4). Those for /g/ were 2395 (F2), 3000 (F3) and 3305 Hz (F4). Stimuli were 220 ms in duration, including 10 ms linear shaping at onset and offset. The VC formant transitions occurred at 60-90 ms, the stop-gap silence at 90-130 ms, and the CV formant transitions at 130-160 ms. There were three varieties of phoneme deviants, where only the VC, only the CV, or both formant-transition regions were changed from /b/ to /g/. The resulting four stimuli (/ibi/, /ibgi/, /igbi/ and /igi/) have important phonetic differences but these phonetic details are not critical here except to indicate that the stimulus set was phonetically complex. For the intensity deviants, the intensity increases were restricted to the formant-transition regions of the VCVs. Three varieties of intensity deviants were used, with intensity increases in only the VC, only the CV, or both formant-transition regions. This variety of intensity deviants (/iBbi/, /ibBi/ and /iBi/) is of no particular phonetic interest, but it also adds to the complexity of the stimulus set. The magnitude of intensity deviation was adjusted in pilot experiments to equate discriminability in intensity and phoneme conditions.

Stimuli in each ear consisted of 70% standards /ibi/, 15% phoneme deviants and 15%
intensity deviants. ISIs varied randomly between 80-125 ms in a rectangular distribution, such that
the average rate of stimulus delivery was 3/s. Subjects attended to a prespecified ear to detect one
category of deviant stimuli (either intensity or phoneme) while ignoring all stimuli in the opposite ear.
Subjects responded with a button press to deviants of the target category in the attended ear. Hand
of response was counterbalanced across the two sessions. Target category and attended ear were
counterbalanced within the 20 recording blocks of each session.

The EEG (bandpass 0.01 to 100 Hz) was recorded from 30 electrodes: BE (below left eye),
LE (lateral to left eye), Nose, Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC6, LM (left mastoid), T7, C3,
Cz, C4, T8, RM (right mastoid), CP5, CP6, P7, P3, Pz, P4, P8, PO5 (center of O1-P3-P7 triangle),
PO6 (center of O2-P4-P8 triangle), O1 and O2. Electrode recordings were referenced to four
interconnected, EKG-balanced electrodes at the base of the neck [13].

Difference waves were obtained by subtracting ERPs to standard VCVs when ignored from
ERPs to the same VCVs when attended during. Nd amplitudes were measured at all electrodes except
for parietal/occipital sites in 20 ms intervals relative to a 200 ms prestimulus baseline. Nd amplitudes
were statistically analyzed with analysis of variance (ANOVA) for discrimination condition (phoneme
vs. intensity); and hemisphere (left vs. right, collapsed across electrode sites within each hemisphere).
Analysis of the Nd scalp distributions included all electrode sites.

**Results**

Reaction times were similar in phoneme (402 ms) and intensity conditions (392 ms)
[F(1,12) = 1.31, ns]. However, hit rates were slightly higher during phoneme detection, 90% vs.
83% [F(1,12) = 10.56, p < 0.01], without a significant change in false alarm rate (less than 1% in
both conditions).

Nd waves from posterior frontal sites over the left and right hemispheres are shown in
Figure 1 for the phoneme and intensity conditions. Overall attention effects reached significance
at 100-120 ms, increased in amplitude until 320 ms, and remained significant throughout the 500
ms duration measurement epoch [F(1,12) ranged from 8.52 (p < 0.02) to 58.28 (p < 0.0005)].
Nd amplitudes were significantly larger during phoneme than intensity conditions. This condition
effect reached significance at 120-140 ms, increased in magnitude before the Nd_t peak, and
remained significant through the 400-420 ms interval [F(1,12) ranged from 5.08 (p < 0.05) to
19.72 (p < 0.002)]. For both conditions Nd_e (110-230 ms) amplitudes were considerably smaller
than Nd₁ (300-500 ms) amplitudes.

The scalp distribution of Ndₑ (measured at 150-200 ms) and Nd₁ (measured at 300-350 ms) are shown in Figure 2 for phoneme and intensity conditions. The Ndₑ had a central scalp distribution whereas the Nd₁ was more frontally distributed. Attention effects showed only a slight trend toward hemispheric asymmetries at short latencies, in both intensity and phoneme conditions, with the asymmetry reaching significance only in the Nd₁ latency range, [280-380 ms, F(1,12) ranged from 4.76 (p < 0.05) to 6.65 (p < 0.05)]. The enhancement of Nd₁ amplitudes over the left hemisphere was more pronounced during phoneme conditions, producing a Condition x Hemisphere interaction [300-360 ms, F(1,12) ranged from 6.53 (p < 0.05) to 13.80 (p < 0.005)]. This interaction is apparent in Figure 2, where the Nd₁ scalp distribution for the phoneme condition shows the clearest left-right asymmetry while the Nd₁ scalp distribution for the intensity condition appears almost symmetrical.

**Discussion**

The results demonstrate that the Nd amplitudes were significantly enhanced when subjects attended to the phonetic rather than the acoustic features in the same sequence of complex speech signals. Since stimulus sequences were identical in the two conditions, these differences cannot reflect acoustic differences between conditions. Nor can they be explained in terms of task difficulty. A number of studies have shown that Nd amplitudes are reduced with reductions in the distinctiveness of attended and nonattended channels [2, 14], but inter-channel discriminability was held constant in the current experiment. And although attention effects generally increase with increasing difficulty of target/standard discrimination [15], this factor would predict an equivalent or larger Nd during intensity than phoneme-discrimination conditions. An alternative explanation is that the magnitude of the Nd reflects the number of neurons engaged in attention-related processing. Insofar as human auditory cortex is specialized for phoneme processing, more neurons might be recruited during phoneme than intensity conditions, producing enhanced Nd components during phoneme attention.

Our results also bear on discussions of hemispheric specialization for the phonetic analysis of speech. A variety of evidence suggests that the left hemisphere may be prepotent in processing stop consonants like those used in the current experiment [2, 16,17,18]. However, the results of
current experiment suggest that the attentional-modulation of phonetic analysis is initially similar in the left and right hemispheres, with asymmetries emerging primarily at longer latencies. Indeed, the long latency of the Nd asymmetries suggest that they may reflect differential hemispheric involvement in the final stages of phoneme feature integration and stimulus identification. Alternatively, the frontal distributions of the asymmetric Nd suggests that it may reflect activity associated with working memory activation, including possible rehearsal in the articulatory-phonological loop[19].
Conclusion

Selective attention effects on auditory processing can be seen in the Nd wave to phonetic stimuli when features attended are embedded in the consonant of a VCV. Attention to the consonant category in the VCV resulted in Nd amplitudes that were significantly enhanced compared to attention to intensity changes in the same portion of the speech signal. Both early and late negative difference waves (Nd_e and Nd_l) were enhanced during phoneme-discrimination conditions. A hemispherically asymmetrical, leftward amplitude asymmetry was evident only during the Nd_l and was more pronounced during phoneme discrimination. The results indicate that auditory cortex of both hemispheres contains specialized mechanisms for phoneme processing, with hemispherically specialized mechanisms engaged primarily during the final stages of phoneme analysis.
Figure Captions

**Figure 1:** Nd waves at posterior frontal sites over the left and right hemispheres (LH and RH). LH averages include electrode sites F7, F3, FC5, T7, C3, CP5, P3 and PO5; RH averages include corresponding right electrode sites. Note the greater Nd amplitude for the phoneme task versus the intensity task and for the LH versus the RH, especially around 300 ms.

**Figure 2:** Scalp distributions at 150-200 ms and 300-350 ms to illustrate the amplitude distributions of Nd\(_e\) and Nd\(_l\) waves, as well as the greater Nd\(_l\) left-right asymmetry in the phoneme task. The same voltage scales were used for phoneme and intensity tasks, but the Nd\(_l\) scale covers a larger range than the Nd\(_e\) scale because the Nd\(_l\) had much greater amplitude. The phoneme-Nd\(_l\) distribution shows the clearest asymmetry, reflecting the task X hemisphere interaction present at this latency.
References


