

Neurophysiological and Behavioral Evidence for a Phonetic Processor

Robert Allen Fox, Jeanne Gokcen and Sheri Wagner¹
The Ohio State University

This session of the CLS meeting represents a phonetics panel which is to discuss, among other things, the proposition that there is a separate speech module specifically responsible for the perception of speech sounds as opposed to non-speech sounds. This phonetic processor would operate in concert with a more general auditory/acoustic processor. Several different versions of this module have been proposed, but the best known is probably that found in Liberman and Mattingly's revised motor theory of speech perception (1985). In their view, the objects of speech perception are the intended phonetic gestures of the speaker and speech perception and production are intimately linked through this specialized module such that the module is responsible for both the perception and production of phonetic structures. The claim of a separate phonetic processor can be seen as only a subpart of the more general claim of the existence of a separate domain-specific language module (Fodor, 1983). Our view of a phonetic processor is that it is separate cognitive mechanism that is specialized for processing speech sounds and making phonetic decisions—it does not include necessarily the rest of the Fodorian canon (e.g., information encapsulation).

A related issue beyond the claim of domain-specificity of the processor, is whether such a special system is innate or is a result of an "innate predisposition" (Kuhl, 1991). In what is certainly to become a very important work on the subject, Elman et al. (1996) have suggested that innateness can be discussed in terms of both mechanisms and content. In terms of mechanisms, Elman and his colleagues suggest that innate constraints could operate at three different levels: representation, architecture, and timing. Representation basically involves knowledge; architecture involves neural structure; and timing involves developmental and maturational changes. Of course, content refers to the nature of the domain—in this case language. As they warn, special content—which many of us would claim for speech—does not necessarily require special mechanisms.

Some arguments supporting a separate speech processor

Several different arguments have been offered to support the claim of a separate speech module, including the following:

- The argument that humans process speech sounds differently from non-speech sounds. Part of the evidence to be discussed here involves a demonstration that humans respond to a set of speech sounds in one way, but non-speech sounds in a different way. For example, several older studies argued that speech sounds were discriminated categorically, while non-speech sounds were discriminated in a more continuous fashion (Liberman et al., 1954, 1956, 1967). In addition, Mattingly et al. (1971) demonstrated that third formant transitions were heard only as tone glides when presented alone but served to make a phonetic distinction when placed in the appropriate syllabic context. Although we now know that non-speech sounds can be identified categorically under certain circumstances (Cutting and Rosner, 1974; Miller et al., 1976) and that discrimination of speech sounds need not be categorical

(Macmillan et al., 1988), these studies served to support the earliest claims of phonetic/auditory processing difference.

- The species-specific argument provides data showing that humans respond to speech sounds in a way differently from non-human species. For example, Kuhl and Miller (1977) demonstrated that the slopes of the identification functions of a synthetic VOT continuum were different in humans than in chinchillas. More recently, Kuhl (1991) presented data indicating that human adults and infants demonstrate a perceptual magnet effect, but Rhesus monkeys do not. These claims are normally based on the assumption that although the mammalian auditory systems of these non-human species and humans are very similar, their responses to speech sounds are significantly different, which indicates the presence of a species-specific phonetic module. However, there are studies which suggest that non-human species can learn phonetic categories (e.g., Kluender et al., 1987).

- Many claims of special speech mechanisms have been supported by brain localization data showing that specific areas of the brain are devoted to the processing of speech based on the assumption that innate systems inherit their own dedicated neural architecture (Fodor, 1983, Pinker, 1994, Elman et al., 1996). Data include both lesion studies and brain activity studies (PET, MRI, ERPs; e.g., Binder et al., 1994; Demonet, et al., 1992; Hari, 1991; Price, et al., 1996; Zatorre, et al., 1991). Evidence for specialized neural architecture for linguistic perception goes back to Broca and Wernicke in the 1860s and 1870s. In the early 1960s, Kimura (1964, 1967) demonstrated that speech sounds were processed primarily in the left hemisphere, while nonspeech sounds were either processed primarily in the right hemisphere or were not lateralized at all. More recently, Creutzfeldt et al. (1989), recording directly from the cortex, found the superior temporal gyrus to be responsive to linguistic stimuli, but not responsive to non-linguistic stimuli. These localization studies can be augmented by lesion studies such as those by Kleist (1934) and Luria (1966) (cited in Creutzfeldt et al., 1989) which have shown that lesion of the superior temporal gyrus on both sides impairs a person's ability to recognize sounds and to distinguish between language and other sounds.

Focus of the current study

The focus of this paper is to examine two different speech perception phenomena, Duplex Perception and the Perceptual Magnet Effect, in terms of both behavioral and neurophysiological data, and to discuss whether these data support the existence of a special phonetic processor. In addition, we will speculate on what the biological origin(s) of such a processor might be, although we feel at the offset that the claim of innate knowledge (e.g., Leslie, 1992; Lightfoot, 1989, and Crain, 1991) is likely to be too strong a claim and simply wrong.

DUPLEX PERCEPTION

The phenomenon of duplex perception, which Doug Whalen addresses in this volume, is one of those remarkable effects that serves as relatively strong support for a distinction between auditory and phonetic processing. The basis of the effect stems from the observation that formant transitions are heard differently, depending on whether they are heard within a syllabic context or in isolation (Mattingly et al., 1991). In isolation these formant transitions sound like nonspeech chirps. However, when they are embedded in synthetic syllables, unique linguistic events

are heard: [ba], [da], and [ga], with no chirp-like sounds heard at all. One can argue that these two percepts are the output of two distinct processors.

However, if the same acoustic signal can provide information to two different modules, the question arises regarding how this information is shared by these two modules. Liberman and Mattingly (1985) contend that the phonetic module and the auditory module compete for acoustic cues. Whalen and Liberman (1987) further suggest that processing for the phonetic module has priority over processing for general auditory perception.

In general, duplex perception can be produced by dividing a CV syllable into two separate parts. One part represents the "syllable base" (including the first two formant transitions and the steady-state portion of the vowel). The second part represents the relatively short third-formant transition which serves as a critical cue to the consonant's place of articulation. The syllable base, heard alone, sounds like a relatively ambiguous syllable (e.g., as either [da] or [ga]), but is heard clearly as a speech sound. The transition, heard alone, is perceived as a chirp. When the two elements are combined under normal intensity levels, they are heard as an unambiguous syllable. However, if the presentation level of the short transition is increased, the listener begins to hear two different percepts: the unambiguous syllable and the chirp. The claim that these two percepts were produced by two different processors is supported by the results of Mann and Liberman (1983) who found that the discrimination functions for these two simultaneous percepts were very similar to those produced separately for (1) the isolated transition, and (2) the whole syllable found by Mattingly et al. (1971), depending on which percept listeners were asked to attend to.

Whalen and Liberman (1987) also used sinusoids to represent the F3 transitions. This was done to reduce the likelihood that any linguistic information could be obtained from the F3 transitions in isolation. These tone glides are composed of a single frequency-changing sinusoidal component unlike formants in natural speech which are complex combinations of different harmonics. Duplexity thresholds (the lowest intensity level of the F3 transition at which both a syllable and tone glide were detected) were determined for both syllables. Both parts of the signal were presented to both ears, and when a token was presented that had the third formant transition intensity increased, listeners simultaneously heard a tone glide and a whole unambiguous syllable.

Presenting all parts equally to both ears avoids some of the problems related to the dichotic presentation of duplex signals which was used in some of the earliest duplex studies. Whalen and Liberman maintain that since this duplex perception occurs when the acoustic input is held constant, it cannot be attributed to any auditory interactions arising from changes in acoustic context or to a shifting of attention between two forms of an ambiguous signal. They argue for the existence of separate auditory and phonetic modes of perception and claim that their results indicate that the phonetic module has priority over the auditory module. Since a tone glide is not perceived until the intensity of the third formant transition is increased relative to the base, they surmise that the incoming signal is processed first in the phonetic module for its linguistic information and that is why a syllable is heard without a tone glide. The remainder is then forwarded to the auditory module for processing, at which point a tone glide is heard.

Although the claims of these duplex perception studies are not without criticism or alternative explanations (e.g., Bailey & Hermann, 1993; Fowler & Rosenbloom, 1992), it seems clear that these behavioral data may provide relatively strong support for the contention that different cognitive subsystems can be made to

respond in different ways to the same acoustic signal. What about neurophysiological data related to this effect?

Neurophysiological implications

If one suggests such a strict separation of perceptual functions (auditory vs. specialized phonetic processing), then it is tempting to assume the existence of some type of biological neurophysiological substrate for such a division, and our study looks for evidence of such a neurophysiological separation. The idea of a separate neurophysiologically based module for specific function is not so far-fetched, as there are examples in the animal kingdom of such specialized modules. For example, it has been suggested that biologically coherent systems underlie echolocation in bats (Suga, 1984) and song in birds (Marler, 1970; Thorpe, 1958). (There is even evidence in humans that such modularity exists in terms of visual depth perception, Julesz, 1960, 1971; Poggio, 1984.)

One of our assumptions is that if a phonetic module exists that is separate from an auditory module in a meaningful and verifiable manner, then it is likely to be instantiated in different patterns of neural activity. Such differences may be found by examining event-related potentials (ERPs) which can provide insight into the underlying neural events that occur as a result of cognitive processing. ERPs have been used to investigate various aspects of speech and language processing, although with somewhat varying results. Some studies have found that linguistic stimuli, as opposed to nonspeech stimuli, evoke different components of potentials (Ratcliff & Greenberg, 1972) while others have noted hemispheric differences based on the linguistic content of the stimuli (Neville, 1974; Hillyard & Woods, 1979; Molfese & Schmidt, 1983). Many studies have shown a difference in latency, polarity, and/or amplitude of different components of an event-related potential, with linguistic stimuli evoking a longer latency or increased amplitude (Wall et al., 1991; Cohn, 1971; Ratcliff & Greenberg, 1972).

Inconsistencies in these studies may be explained by several factors. First, most of these studies did not control adequately for noncerebral artifacts such as eye movement which could have contaminated the data. Second, a problem with most (if not all) of these early studies is that the speech and nonspeech stimuli were not matched in terms of such acoustic characteristics as frequency, duration, number of formants, nature of frequency transitions, etc. This makes it more difficult to ascertain differences on the basis of linguistic characteristics of the speech stimuli vs. the nonspeech stimuli, or to the multiple acoustic differences which separate the two classes of stimuli. However, the nature of the stimuli used in duplex perception should avoid many of these problems because they are matched in terms of both the number and nature of their components.

Evoked potential experiment

In the ERP study to be described (derived from Gokcen, 1992), brain ERPs were recorded while subjects listened to and silently identified plain speech-only tokens, duplex tokens, and tone glides. The dependent variables were the amplitude of the maximum excursion (peak) in microvolts and the time of the maximum excursion in milliseconds relative to the onset of the stimulus, also known as the "latency" of that peak. Several different questions were addressed:

1. Is there a significant difference in the ERP waveforms of speech-only, nonspeech-only, and duplex stimuli in terms of their latencies?

- A. Do listeners show shorter latencies or lower amplitudes to speech-only tokens than to duplex tokens (e.g., [da] vs. [da] + tone glide; [ga] vs. [ga] + tone glide)?
 - B. Is there a significant effect of the consonant differences alone on ERP latencies or amplitudes?
 - C. Are listener latencies or latencies significantly affected by direction of frequency change when listening to the F3 transitions alone?
2. Is there a significant difference in the ERP waveforms of speech and nonspeech stimuli in terms of their base-to-peak amplitudes?

Assuming the existence of a specialized phonetic module and assuming that the operations of this module are neurophysiologically differentiable from those of general auditory processing, then we might expect that the ERP latencies recorded while listening to and identifying duplex tokens may be significantly longer than latencies associated with speech-only (plain) tokens or the tone glides.

Methodology

Stimuli. Stimuli were based on the tokens in Whalen and Liberman (1987) and represent synthesized versions of the consonant-vowel (CV) tokens [da] and [ga] 250 msec in total duration. The CV base, synthesized using the cascade version of the Klatt synthesizer (Klatt, 1980) consisted of the first two formant transitions, 50 msec in duration, and the steady-state portion of the all three formants 200 msec in duration. The frequency of the F1 transition began at 279 Hz and changed to vowel's steady-state value of 765 Hz. The F2 transition started at 1650 Hz and fell to the steady-state value of 1230 Hz. The steady-state frequency of F3 for the vowel was 2527 Hz. There was no stop release burst.

The two tone glides representing the F3 transitions for [da] and [ga] were 50 msec in duration. The [da] tone glide transition changed from 2800 at token onset to 2527 Hz at vowel onset. The [ga] tone glide transition changed from 2018 Hz at token onset to 2527 Hz at vowel onset. Fundamental frequency (F0) of the base began at 120 Hz and dropped linearly to 110 Hz by the end of the token, which produced a more natural sounding speech token. The base and one of the transitions were output directly from the computer to headphones in two separate channels in a time-synchronous manner to produce a single stimulus token. D-to-A conversion used 12-bit quantization, a 10 kHz sampling rate, and a low-pass smoothing filter at 4.5 kHz. Subjects heard the tokens binaurally with an interstimulus interval (ISI) of 1646 msec. Output level was calibrated at 75 dB.

The base, when presented in isolation, was heard as an ambiguous CV syllable, mostly as [da]. When the transitions were presented in isolation, they were heard as nonspeech tone glides. When one of the transitions was combined with the base at an amplitude equal to the base, the stimulus token was heard unambiguously as the CV syllable corresponding to the slope of the F3 transition (i.e., as either the [da] or [ga]). When the level of the F3 transition was increased by 6 dB, both the appropriate CV syllable and a tone glide were heard, simultaneously.

Electrophysiologic activity from subjects was monitored while subjects listened and silently identified each token as it occurred. There were three types of comparisons which were subdivided into five conditions:

1. [da] vs. [da] + tone glide
2. [ga] vs. [ga] + tone glide

3. [da] vs. [ga]
4. [da] + tone glide vs. [ga] + tone glide
5. [da] tone glide vs. [ga] tone glide

In Conditions 1 and 2, the two stimulus tokens presented differed only in terms of the presence or absence of the duplex "chirp." In Conditions 3 and 4, the stimulus tokens differed in terms of initial consonant only, and in Condition 5, the stimulus tokens differed in terms of the onset and slope of the tone glide. These five conditions were presented to each subject in the order listed above to reduce subjects' exposure to the tone glides as nonspeech tokens prior to their exposure to the tone glides within the context of speech. Within each condition, there were 50 occurrences of the two token types, for a total of 100 tokens presented in pseudorandom order.

In order to control the subject's attention and involvement during testing, a simple behavioral task was included while the ERPs were being recorded. Subjects were to silently identify each token as it occurred and there was no single target that subjects were listening for. Prior to testing, subjects participated in a training session. To minimize eye movement, subjects were asked to focus on a dot placed on a flat surface at eye level. A mean of 32 trials out of 50 per subject were averaged to obtain waveforms, with a range of 10 to 49 trials.

Brain electrophysiological responses were obtained from fifteen active gold cup electrodes placed on a subject's scalp, following the International 10-20 convention (Jasper, 1958; at Fz, Cz, Pz, F3, C3, P3, F4, C4, P4, F7, T3, T5, F8, T4, and T6) each referred to linked ears. Fpz on the forehead served as a ground. Eye movements and blinks were monitored via electrodes placed at the right external canthus and above the left supraorbital ridge. The electrodes were connected to a 20-channel electrophysiological data acquisition and analysis system (Brain Atlas, Biologic Systems Corp.). Subjects served as their own controls.

For the present study, only one recording channel was analyzed, Cz, because typically it has a high signal-to-noise ratio and is one of the most often used ERP site in the literature. All relevant waveforms were identified, noting polarity, and measuring baseline-to-peak amplitudes and peak latencies. ANOVAs were performed on these data within each condition.

Results

This experiment produced much more data than can easily be presented within the current context, but we will address what we consider the most salient points. Latency and amplitude measurements were made at three points in the waveform: N1, P1, and N2 which refer to the first negative peak, the first positive peak, and the second negative peak², respectively.

As noted earlier, in both Conditions 1 and 2, listeners heard a plain speech-only token with its duplex counterpart. Mean latencies and amplitudes are shown in Table 1. In general, the duplex stimuli for both [da] and [ga] produced significantly longer latencies than did the speech-only tokens. Two-way repeated measures ANOVAs with the factors syllable type (speech-only and duplex) and syllable ([da] and [ga]) showed significant main effects of syllable type for N1 ($F(1,15)=253$, $p<.001$), P1 ($F(1,16)=119$, $p<.001$), and N2 ($F(1,14)=15.4$, $p<.001$). The main effect of syllable and the syllable type by syllable interaction were not statistically significant for any of these three measures. The longer latencies that were found for the duplex stimuli suggest that additional neural processing of some kind may be involved when both a speech and non-speech percept is produced.

In general, there were relatively small effects on amplitude values in Conditions 1 and 2 (Table 1) as a function of the syllable type or syllable. Two-way repeated-measures ANOVAs of the amplitude data showed a significant effect of syllable type for N1 ($F(1,13)=5.1, p<.05$) and P1 ($F(1,15)=5.3, p<.04$) but not for N2. Only P1 showed a significant main effect for syllable ($F(1,15)=21.9, p<.001$) and none of the syllable type by syllable interactions were significant.

Table 1. Mean latencies and amplitudes for Conditions 1 and 2 (standard deviations in parentheses).

	Syllable	Syllable Type				Mean	
		Plain		Duplex			
N1 Latencies	[da]	202	(10.5)	234	(9.2)	218	(9.9)
	[ga]	204	(11.5)	234	(8.3)	219	(10.1)
	Mean	203	(11.0)	234	(8.8)		
P1 Latencies	[da]	294	(10.6)	319	(11.6)	307	(11.1)
	[ga]	292	(12.7)	319	(9.3)	307	(11.0)
	Mean	293	(11.7)	319	(10.5)		
N2 Latencies	[da]	393	(29.3)	404	(20.4)	399	(24.9)
	[ga]	378	(16.1)	407	(19.9)	393	(18.0)
	Mean	388	(22.7)	405	(20.2)		
N1 Amplitude	[da]	-11.7	(4.6)	-13.8	(4.7)	-12.8	(4.7)
	[ga]	-11.2	(4.9)	-12.6	(5.5)	-11.9	(5.2)
	Mean	-11.5	(4.8)	-13.2	(5.1)		
P1 Amplitude	[da]	12.2	(3.7)	15.0	(6.4)	13.6	(5.1)
	[ga]	9.8	(3.4)	10.4	(2.9)	10.1	(3.2)
	Mean	11.0	(3.6)	12.7	(4.7)		
N2 Amplitude	[da]	-7.0	(4.3)	-5.8	(4.3)	-6.4	(4.3)
	[ga]	-7.1	(3.8)	-5.7	(6.4)	-6.4	(3.4)
	Mean	-7.1	(4.1)	-5.8	(3.7)		

Measures N1 and P1 are early components in the neural response and may be a reflection of the subject's attention to the stimuli (Picton & Stuss, 1984). The increased amplitude of the F3 transition and the duplex nature of the percepts may have made the duplex tokens more noticeable to the subjects.

In Conditions 3 and 4, subjects heard either the two speech-only tokens or the two duplex tokens and the difference between the two tokens in each condition involved a linguistic difference only ([da] vs. [ga]). The mean latencies and amplitudes obtained in these conditions are shown in Table 2. The latencies of N1, P1, and N2 were again separately analyzed using two-way repeated measures ANOVAs with the factors syllable type and syllable. There was a significant main effect of syllable type for N2 ($F(1,7)=7.3, p<.03$) but not for N1 or P1. Significant main effects of syllable were found for N1 ($F(1,11)=118, p<.001$), P1 ($F(1,11)=168, p<.001$) and N2 ($F(1,17)=7.32, p<.03$). None of the syllable type by syllable interaction effects were significant.

Table 2. Mean latencies and amplitudes for Conditions 3 and 4 (standard deviations in parentheses).

		Syllable Type					
	Syllable	Plain		Duplex		Mean	
N1 Latencies	[da]	198	(8.4)	201	(10.7)	200	(9.6)
	[ga]	231	(11.1)	234	(11.0)	233	(11.2)
	Mean	215	(9.8)	218	(10.9)		
P1 Latencies	[da]	288	(10.0)	290	(13.8)	289	(11.9)
	[ga]	316	(12.9)	319	(8.6)	318	(10.5)
	Mean	302	(11.5)	305	(11.2)		
N2 Latencies	[da]	372	(23.0)	383	(17.1)	378	(20.1)
	[ga]	403	(25.0)	441	(56.6)	422	(40.8)
	Mean	388	(24.0)	412	(36.9)		
N1 Amplitude	[da]	-13.5	(5.0)	-10.6	(4.0)	-12.5	(4.5)
	[ga]	-11.5	(5.1)	-9.0	(4.1)	-10.3	(4.6)
	Mean	-12.5	(5.1)	-9.8	(4.1)		
P1 Amplitude	[da]	11.4	(4.1)	11.3	(3.3)	11.4	(3.7)
	[ga]	11.2	(4.9)	11.0	(3.7)	11.1	(4.3)
	Mean	11.3	(4.5)	11.2	(3.5)		
N2 Amplitude	[da]	-5.4	(3.4)	-6.0	(3.3)	-5.7	(3.4)
	[ga]	-7.1	(1.9)	-6.0	(2.2)	-6.6	(2.1)
	Mean	-6.3	(2.7)	-6.0	(2.8)		

Table 3. Mean latencies and amplitudes for Condition 5 (standard deviations in parentheses).

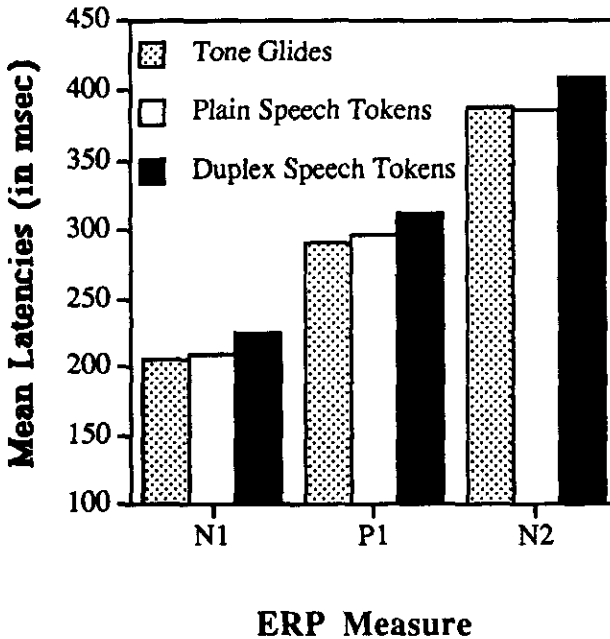
		Tone Glide Data					
	Measure	[da] Transition		[ga] Transition		Mean	
Latency	N1	189	(10.1)	222	(11.4)	206	(10.8)
	P1	273	(20.1)	308	(16.7)	291	(18.4)
	N2	368	(15.1)	407	(14.1)	388	(14.6)
Amplitude	N1	-11.8	(2.9)	-9.2	(4.0)	-10.5	(3.5)
	P1	13.5	(7.0)	11.3	(6.3)	12.4	(6.7)
	N2	-3.9	(4.5)	-4.7	(2.7)	-4.3	(3.6)

Latency values were consistently longer for [ga] than [da]. This result may indicate that [ga] is less perceptually salient than is [da], perhaps because we are using synthetic speech without any release bursts, each of which will reduce the naturalness of the tokens. In fact, the majority of subjects reported more difficulty in identifying [ga] than [da].

Analysis of the amplitude measures indicated significant main effects of syllable type ($F(1,10)=10.8$, $p<.01$) and syllable ($F(1,10)=7.3$, $p<.03$) for N1 only. None of the syllable-type by syllable interactions were significant. As in

Conditions 1 and 2, differences in amplitude values were mostly not statistically significant. Larger amplitude values of N1 for [da] compared to [ga] may again reflect the decreased salience of [ga].

Figure 1. Mean latencies obtained for N1, P1, and N2 across all five conditions.



In Condition 5, the two isolated tone glides/transitions were presented. This may not necessarily involve a linguistic difference, but it is outside any speech context. The latencies and amplitude values obtained are presented in Table 3. One-way ANOVAs with the within-subject factor tone contour (corresponding to the falling [da] or rising [ga] transition) showed significant main effects on latency for N1 ($F(1,12)=374, p<.001$), P1 ($F(1,14)=64, p<.001$) and N2 ($F(1,10)=45.2, p<.001$). For amplitude values, there was a significant main effect for P1 ($F(1,14)=4.52, p<.052$) but not for N1 or N2.

A tone contour effect is not unexpected as these two tokens are auditorally quite distinct. The increased latency for the [ga] transition is probably due to its greater extent of frequency change over time compared with [da].

If there are two separate perceptual components, then we might expect that the latency values for the duplex token would be greater than for the speech-only tokens or for the tone-glide only tokens. Shown in Figure 1 are the mean latencies obtained in all five conditions. As can be seen, the duplex tokens show longer mean latencies for all three ERP measures. One might predict that when subjects are forced to compare two tokens within a single condition that differ in terms of both a linguistic distinction and a non-speech distinction (e.g., speech-only [ba] and duplex [ga]), that even longer latencies would occur. Unfortunately, such a

comparison was not included in the present study, but will be examined in later experiments.

Discussion

There was considerable variability across subjects, especially in the amplitude values. Latency values were typically more consistent across subjects. Results indicated the neural activity in response to the stimuli was different for the stimulus types. Duplex tokens had significantly longer latencies than the pure speech tokens. Amplitude differences were mostly insignificant.

Results did demonstrate that the neural activity in response to the stimuli, as measured by the peak latencies, was different for the two stimulus types. When plain speech-only tokens were compared with duplex tokens, the duplex tokens produced longer latencies than did the plain speech-only. Even in Conditions 3 and 4, although the syllable type difference was generally not significant, the mean latencies for the duplex tokens were longer (see Table 2). These results appear to support Whalen and Liberman's (1987) hypothesis that the phonetic module first utilizes the information it needs to identify the speech component and after that processing is completed, the remaining "extra" nonspeech information is then processed by the general auditory module. Subjects did not report hearing a tone glide, that is, they heard only a CV syllable, until the tone glide's amplitude level was increased.

Even if the explanation for the latency differences is simply because linguistic and nonlinguistic components have two different areas in the brain to which they must go for processing, and coordinating these two processing sources in order to make an identification of a stimulus takes longer, the data would be consistent with the contention of separate modules for phonetic and auditory stimuli. We would argue that these data do not support the claim that there is only a single unified cognitive module that processes all auditory information because the speech-only and duplex stimuli contained identical components and were equal in complexity³.

When duplex tokens are compared with each other and plain speech-only tokens are compared with each other, respectively, the latency difference are a reflection of the phonetically different initial consonants. In Conditions 3 and 4, the results indicate that there generally was no overall main effect of stimulus type, but a strong effect of syllable; [ga] produced longer latencies than did [da] regardless of whether it was the plain or duplex [ga].

When the two tone glides alone were presented there was a significant latency effect related to the different tone contours. Subjects were able to differentiate one tone glide from the other, although they did not consistently identify either tone glide as being associated with their respective consonants (in the test they were to try to identify these as [ga] or [da]). It is unclear why the latencies would be longer for the [ga] tone glide, although we note the difference is in the same direction as in Conditions 3 and 4.

The results can also be viewed in regard to the task requirements, which were different among the conditions. In Conditions 1 and 2, subjects had a largely nonlinguistic task: to identify tokens which either did nor did not contain a "chirp." In Conditions 3 and 4, subjects had a largely linguistic task: to identify [da] or [ga] tokens. In Condition 5, subjects also had a linguistic task: to identify the tone glides as either a [da] or [ga]. These task differences were reflected in the results: for Conditions 1 and 2, the factor stimulus type was significant in terms of latency

values for all ERP components and syllable was insignificant, while the opposite was true for Conditions 3, 4, and 5.

Amplitude differences did not show any consistent and significant pattern and was highly variable among subjects. From the amplitude data, there is no indication that the processing of one type of auditory stimulus, be it speech or nonspeech, involves any more or less neural activity than another.

Future research will concentrate on analyzing other electrode placements to gain information regarding the topographical distribution of the two types of stimuli. Such an analysis should provide insight as to the locations of the two modules, or at least, information about the distribution of neural activity.

THE PERCEPTUAL MAGNET EFFECT

The second phenomenon to be examined is the so-called "perceptual magnet effect." Patricia Kuhl and her colleagues have published several different studies (Kuhl 1991; Kuhl, et al., 1992; Iverson and Kuhl, 1995) examining the internal structure of phonetic categories and have obtained data which demonstrate that category goodness of a speech sound influences the perception of that speech sound. In particular, they suggest that listeners "exhibit relatively poor discrimination in the region of prototypic exemplars of phonetic categories," whereas discrimination performance for stimuli that are not close to a prototype are correspondingly better. The term "perceptual magnet" was coined for this effect since the prototype draws exemplars closer to it in the perceptual space.

Kuhl has made the relatively strong claim that the internal organization of phonetic categories "is an ontogenetically early, species-specific, aspect of the speech code" which, if true, could be a product of the separate phonetic module. Indeed, as noted earlier, Kuhl (1991) has claimed that although human adults and children show this effect, a closely related species, Rhesus monkeys, do not. The perceptual magnet effect is influenced by early exposure to a language (Kuhl et al., 1992) and thus is native-language specific. Clearly, this means that it does not correspond to innate knowledge, but it could stem from innate neural architecture and predispositions (Kuhl, 1993).

However, several concerns have been raised about the stimuli used in Kuhl's studies. In the original study, all the stimulus tokens were considered to represent different versions of a single vowel category /i/, differing only in terms of category goodness. However, as Sussman and Lauckner-Morano (1995) demonstrated, many of these vowels could be perceived by naive listeners as members of a different phonemic category. Thus the effect might be a between-category effect rather than a within-category effect that reflects the internal structure of that category. A related concern for both Kuhl and her colleagues as well as Sussman & Lauckner-Morano is that in studies comparing the discrimination of tokens near the prototype (P) with those near a non-prototypic token (NP), the NP token is always close to (if not over) the category boundary of another phoneme ([e] or [i]), which again muddies the water in terms of whether this is truly a reflection of the internal structure of the category, or a category boundary effect.

In this next section of this paper, we will present the results of a study just recently completed that looks for a perceptual magnet effect when (1) all tokens are clearly representative of a single phonemic category, and (2) the non-prototypic end of the vowel continuum created is not in the direction of another vowel category. Similar to Iverson & Kuhl (1995), the evaluation of whether or not a perceptual magnet effect is present will be based on whether vowels near a prototypic vowel

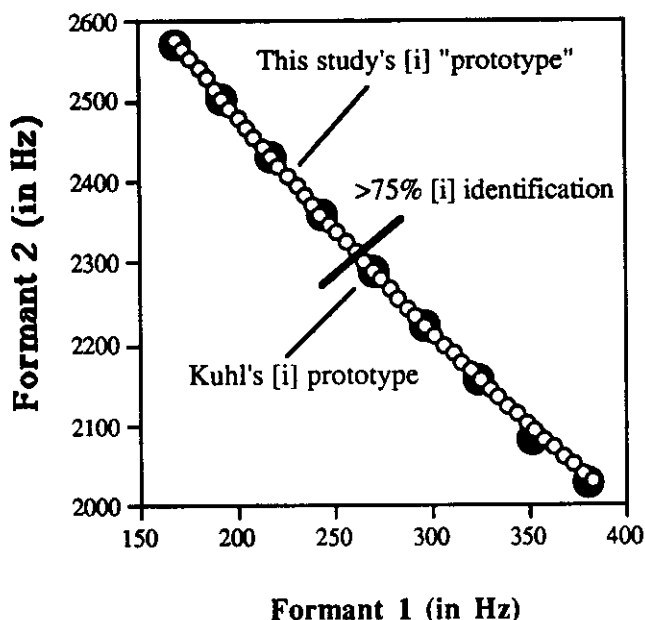
are more tightly clustered in the perceptual vowel space (to be obtained using multidimensional scaling) than are vowels farther away from the prototype.

Following the presentation of these data, we will discuss neurophysiological data recently obtained by Aaltonen et al. (1997) which examine changes in mismatch negativity as a function of the perceptual magnet effect.

Methodology

Stimulus Creation. The synthetic vowel stimuli created for this experiment were based on the original set of vowels described in Kuhl (1991), subsets of which were used in Kuhl et al. (1992), Iverson & Kuhl (1995), and Sussman and Lauckner-Morano (1995). The original vowels represented points in the F1-F2 space arranged in four orbits and eight vectors around a center stimulus. These vowels were spaced along the F1 and F2 axes in 30 mel steps. However, in our experiment (as in Iverson & Kuhl) we are only interested in the vowels along two of the vectors (vectors 4 and 8) as shown in Figure 2.

Figure 2. Illustration of vowel continuum created (open circles); closed circles indicate vowel tokens included in Kuhl (1991), Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995).

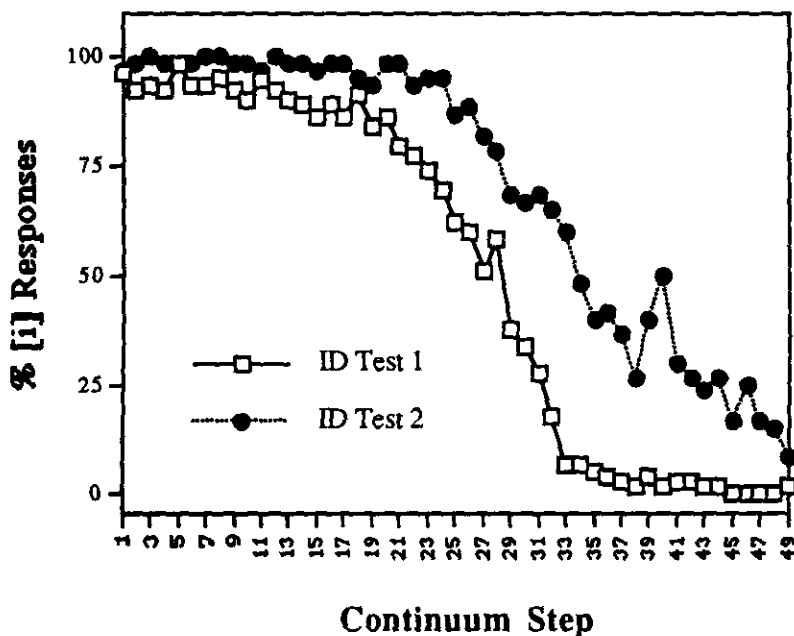


Using the two vowels in the outside orbit of these vectors as endpoints, a vowel continuum having 49 different steps was created. The frequencies of formants 1 and 2 first endpoint were 169 and 2675 Hz; while those of the second endpoint were 380 and 2027 Hz. The vowels within the continuum changed in approximately 5 mel steps for both formants 1 and 2. Formants 3, 4, and 5 remained the same for all tokens at 3010, 3300, and 3850 Hz, respectively. The bandwidths of these five formants were 50, 70, 110, 250, and 200 Hz. The tokens were normalized in amplitude and the onsets and offsets were ramped from or to

zero over 10 msec. In all testing procedures described below, all vowels were presented to listeners at 74 dB SPL (A), using direct digital-to-analog conversion at a 10 kHz sampling rate through a 4.5 kHz low-pass smoothing filter.

Vowel Identification. Given our goal, it was of particular interest that all vowels used in the multidimensional scaling experiment be clearly identified as exemplars of the /i/ phoneme. Therefore, two identification tests were conducted. In the first test, ten listeners heard ten examples of each step of the vowel continuum (490 tokens in all) which were presented in randomized order. Listeners were required to indicate whether the token was an example of /i/, /ɪ/, /e/, or some other vowel. However, it may be the case that listeners magnify small differences when the acoustic variations among the vowels are small, an effect which could lower the number of /i/ responses. For example, one could argue that listeners can be affected by response bias, attempting to even out the number times each response category was used. Thus, in a second identification experiment, ten tokens of clear exemplars of the /e/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/, and /ɑ/ vowels were included with the 49 steps of the continuum; again the order of presentation was randomized.

Figure 3. Percentage of vowel tokens identified as [i]; Test 1 contained only [i]-like vowels, Test 2 contained clear examples of other phonetic categories.

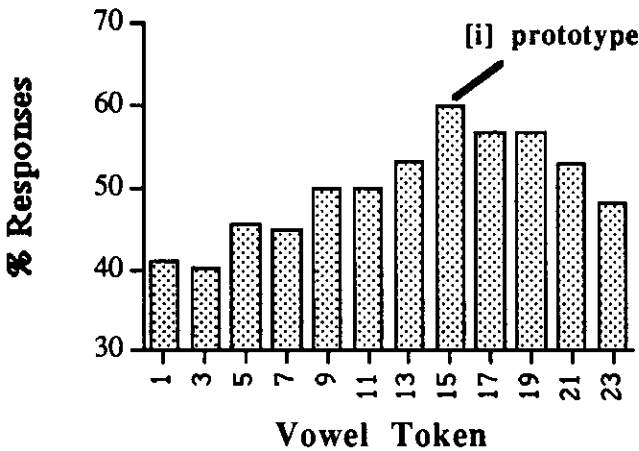


The results of these identification tests are shown in Figure 3. In order to be considered a clear example of the /i/ category, a stimulus token had to be identified as /i/ at least 75% of the time (a criterion consistent with Sussman and Lauckner-Morano). Two things are clear: First, addition of the clear non-/i/ exemplars did affect the identification of relatively ambiguous tokens near the category boundary. Second, as demonstrated by Sussman and Lauckner-Morano,

many of these tokens do not, in fact, represent exemplars of the /i/ vowel category in either identification test. Kuhl's (1991) prototypic vowel, in the first identification task, was identified as /i/ only 62.5% of the time, whereas the endpoint of vector 4 was identified as /i/ only 1.3% of the time.

In all, twenty-three vowels of the continuum met our identification criterion. We chose 12 of these vowels for use in the MDS study (vowel 1, the endpoint stimuli in vector 8, and vowels 3, 5, 7, ... 23). Vowel 23 is only two 5-mel steps away from Kuhl's original prototype vowel (as shown in as shown in Figure 2).

Figure 4. Prototype ("most typical") responses to vowel tokens.



Prototype Identification. Given that all 12 stimuli selected are clearly exemplars of the /i/ category, we needed to identify which vowel token was the "best" or "most prototypic" exemplar of the American English vowel [i]. To do this, we presented to twelve listeners all possible pairs of the 12 stimulus vowels (including identical vowel pairs) in both possible orders (AB and BA) with a 500 ms interpair interval. The presentation order of these vowel pairs was randomized. Listeners were required to indicate whether the first vowel or the second vowel of each pair represented the "most typical" example of the American English /i/. This task not only demonstrates which vowel is most representative of the underlying /i/ prototype, but will provide a measure of the discriminability of the vowel pairs. Shown in Figure 4 are the total number of times each token was identified as the best example of the category, summed across all vowel pairs (except for the identical vowel pairs). As can be seen, vowel 15 (remember only odd-numbered steps along the vowel continuum were used) received the greatest number of responses. Note that the vowels to the left received significantly fewer responses but these non-prototypic vowels were not close to any other American English vowel category).

Multidimensional Scaling. Multidimensional Scaling (MDS) is an experimental technique which examines obtained perceptual distances among stimuli and develops an n -dimensional model which accounts for those distances (e.g., Singh & Woods, 1971; Fox 1982, 1983, 1985; Kewley-Port and Atal, 1989). It is generally accepted that this technique has been effective in modeling the perceptual

dissimilarity among vowels and was used by Iverson and Kuhl (1995) to model reaction times (RT) of discrimination responses to exemplars of the /i/ category (but see concern noted above). Their results indicated that the resulting perceptual space was distorted in the sense that it was "shrunk" in the region near the best instances of /i/ (i.e., near the prototype) and "stretched" in the region of the poorest exemplars of the category. However, their stimuli included exemplars of other vowel categories and thus may not reflect a pure within-category effect nor accurately reflect the internal structure of a single vowel category. Our goal is to determine if this distortion of the perceptual space will occur within our more limited, within-category stimulus set.

Twelve listeners (the same subjects as in the prototype identification experiment) listened to two examples of all possible pairs of the 12 stimulus vowels (including identical vowel pairs) in both possible orders (AB and BA) with a 500 msec interpair interval. The presentation order of these vowel pairs was randomized. Listeners were required to indicate how similar or different the two vowels were on a 9-point scale as follows:

S D

Listeners were warned that the vowels might sound very similar to one another and that the differences between them might be subtle. They were also told that a small number of identical vowel pairs were being presented and were asked to use the entire 9-point scale.

If listeners are inconsistent in their responses, then "noise" will be introduced into the perceptual distances which may distort the obtained perceptual space. To determine how consistent listeners were, each listener's summed responses to all stimulus pairs in one order (A B) were compared to the summed responses in the reverse order (B A) using Pearson's product-moment correlation. All listeners showed a significant correlation coefficient (average $r = 0.711$) and the data from all 12 listeners were used in the MDS analysis.

The rating data from each listener were collapsed into symmetrical vowel-by-vowel matrices so that each cell of a matrix represented the sum of the four responses to a given vowel pair, ignoring presentation order. Note that since these ratings represent a measure of perceptual distance, that it is assumed that the diagonal values should be 0 since they represent the perceptual distance between identical vowels. However, as in Fox, Flege & Munro (1995)—and unlike many other MDS studies—in the current study, listeners rated all possible vowel pairs, including identical pairs. Thus, the mean values in the diagonal following symmetricization were not zero, but were 4.0 or higher. Given the assumption of MDS analysis that the distance from an object and itself is 0, a value of 4.0 was subtracted from each cell, and the diagonal values were set to 0. This procedure has the advantage that vowel pairs that are extremely similar could also receive a value near 0, which may be an appropriate estimate of their true perceptual distance.

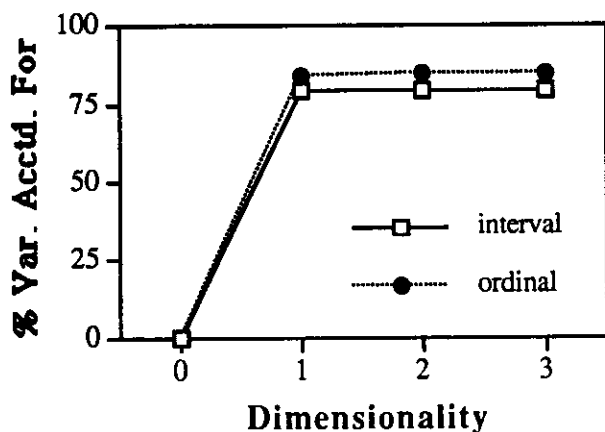
MDS Analysis. The summed ratings were then examined using MDS analysis. This analytic technique accounts for the perceptual distances between pairs of stimuli by locating the stimuli within an n -dimensional perceptual space. Each listener's symmetricized matrices were analyzed using ALSCAL (Takane, Young, and DeLeeuw, 1976). In particular, the weighted individual differences scaling procedure was used, except for the one-dimensional solutions which used the default EUCLID model. No a priori assumptions concerning the number or the nature of the perceptual dimensions underlying the listeners' dissimilarity ratings were made except that they represented a Euclidean space. The data were analyzed

to produce both ordinal (nonmetric) and interval (metric) solutions in 1-3 dimensions.

Selection of dimensionality. The choice of the optimal MDS solution was based on a consideration of the amount of variance accounted for by various solutions. Given the nature of the vowels chosen, the results of Iverson & Kuhl's (1995) study, and the linear acoustic manipulations done to create the vowel continuum (i.e., F1 changes are correlated with F2 changes), we expected the appropriate perceptual space to be one dimension.

Shown in Figure 5 are the fit curves showing the cumulative variance accounted for by both the interval and ordinal solutions as a function of the number of perceptual dimensions. Two things are clear from this figure: (1) The ordinal solutions account for a greater proportion of the variance than do the interval solutions for all five groups at each dimensionality, and (2) the one-dimensional solution seems to be the appropriate one to examine, given the very sharp "elbow" in the fit curve (i.e., a radical change in the slope of the variance-accounted-for curve which indicates that the appropriate dimensionality has been reached).

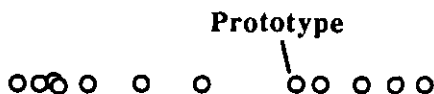
Figure 5. Fit Curves for metric (interval) and nonmetric (ordinal) solutions in 1-3 dimensions.



Discussion

The critical question to be addressed is whether there is a distortion of the perceptual space near what we have identified as the "most typical" exemplar of the [i] phonetic category. Remember that in the magnet effect shown in the Iverson & Kuhl, the vowel tokens near the prototypic [i] (which would have included all the tokens used in our experiment) clustered closely together apart from the nonprototypic [i] tokens were, in turn, were more distant from those tokens that were exemplars of the category [e]. Shown in Figure 6 is the one-dimension solution for the perceptual distances obtained in the present study using clearly within-category comparisons. An arrow in the figure identifies the token that was identified as the "most typical" exemplar of American English [i] for these subjects.

Figure 6. One-dimension MDS solution.



In terms of interpreting the dimension, the positions of the tokens along the coordinate axis related in a straightforward manner to the original acoustic differences among the stimuli. However, there is no evidence of clustering near the "prototypic" vowel. Rather, if any clustering is apparent, it is among those tokens which are in the more extreme range of [i] tokens or in the hyperarticulated perceptual space that Keith Johnson (personal communication) has found were often identified as "prototypes" in his perceptual studies. These data do not provide strong support for the magnet effect demonstrating the internal phonetic category structure as defined by Iverson & Kuhl (1995) in that the vowel tokens that show some clustering are almost never found in normal productions of [i] in the subjects' midwestern dialect of American English and could never serve as representatives of their commonly produced versions of [i].

In terms of our more general interests in this paper, do these results provide counterevidence for the claim for a separate phonetic module? We would argue that they do not. In fact, one can claim that a separate phonetic module would be most sensitive to the difference between tokens that are more representative of one category as opposed to another. It should be *less* able to differentiate among tokens that are all relatively good exemplars of a given category, especially when the most extreme differences are not close to a separate phonemic category (e.g., /i/ or /e/). We would predict that if one placed several tokens that were good exemplars of a different category with the stimuli used here, that significant distortion, of the type and degree found by Iverson & Kuhl, would be obtained.

Neurophysiological implications

Although our investigation of the perceptual magnet effect has involved the collection of behavioral data only, several recently published studies have addressed how this effect is reflected in neurophysiological activity, especially in terms of the measure MMN. MMN refers to mismatch negativity and is an event-related potential which is sensitive to relatively fine acoustic distinctions. It is said to be "preattentive" which means that it may allow an investigator to examine responses to acoustic differences in both speech and non-speech sounds prior to a subject's conscious processing of the stimuli (Sharma et al., 1993). Such a sensitive measure may be able to identify any electrophysiological differences that may underlie the perceptual space distortions discussed previously.

Sharma and his colleagues (Sharma et al., 1993) used MMN to examine within- and across-category distinctions of a /da/-/ga/ continuum in which each step represented equal acoustic differences, but with a clear perceptual category boundary. They found that the MMN did not differ in latency, amplitude or area within and across categories. This was in partial disagreement with Sams et al. (1990) who examined MMN responses to a /bæ/-/dæ/-/gæ/ continuum and found smaller MMN amplitudes for within-category stimuli than to across-category stimuli. Sams suggested that the human auditory cortex could contain neural networks that are activated by speech-specific features such as formant transitions, whether produced by the attraction of a prototype or the presence of a phoneme boundary.

Olli Aaltonen and colleagues have also used MMN to examine acoustic vs. phonetic processing. Aaltonen et al. (1993) recorded event-related potentials from four aphasic subjects and found that all subjects produced MMN responses to sinewave tokens, but two patients with left posterior brain lesions did not show an MMN to synthetic vowels, supporting the conclusion of Sams et al. (1990). In a more recent study, Aaltonen et al. (1997) examined the perceptual magnet effect using MMN to provide evidence on the traditional division of the perceptual process into distinct auditory and phonetic stages. They assumed that looking for the perceptual "magnet effect in subjects who are ignoring the stimuli should ... help to clarify the role of the auditory component in speech perception." Aaltonen et al. conducted a series of four experiments with the same Finnish-speaking subjects using a synthetic vowel continuum differing only in F2 frequency. Both prototype (P) and nonprototype (NP) tokens were identified and used as referents in an AX discrimination task looking at preattentive auditory discriminations and measuring MMN. Their results indicated that one group of their subjects (the Low P group) demonstrated a perceptual magnet effect in both behavioral responses and in terms of MMN amplitudes. They concluded that MMN may be modified by the listener's "experience with spoken language" and that on the "auditory processing level this experience is organized into clusters of similar, frequently heard speech sounds."

CONCLUSIONS AND SPECULATION

What can we conclude and/or speculate from both the data presented above and previous research in terms of the questions: "Is speech special?" and "Is there a difference between phonetic and acoustic processing?"

In terms of the first question, speech is certainly special at least in the obvious sense that only humans naturally use speech or respond differentially to speech as opposed to nonspeech sounds. It is almost certainly true that researchers can train members of nonhuman species to produce responses that appear similar to those that could be obtained from humans. Does this mean that speech is not special to humans? No, because it is clear that similar results do not imply similar means of processing. The neural architecture in the two species could be completely different and yet produce the behavioral responses that are highly correlated. You can program a computer to respond to certain speech sounds in a manner practically indistinguishable from humans—but obviously there is no necessary similarity in the way the acoustic signals are processed or in the decision-making procedure.

Most of the neurophysiological data presented or discussed here suggests that humans display different patterns and/or locations of neural activity to speech sounds as opposed to nonspeech sounds, which would be expected if there is a substantive and meaningful difference between phonetic and acoustic processing. Furthermore, we would argue that the ERP results obtained in the duplex perception experiment in which the acoustic elements of the speech and nonspeech sounds are matched in terms of the number and frequency of the components is consistent with the claim that speech and nonspeech are processed in a differential fashion. In terms of speculation, we would argue that the differences may be the result of species-specific neural architecture. Furthermore, the language-specific response differences found by Kuhl and Aaltonen and others may reflect the nonuniformities in the distribution of preference of neural firings that are a product of the specific sounds experienced by an infant in the first six months of life (and later) rather than any innate knowledge or hardwired constraints related to phonetic properties, *per se*. We find the current connectionist research on this topic

interesting and potentially very insightful (see related discussion by Guenther, 1995, Guenther & Gjaja, 1996; Grossberg, 1976; Kohonen, 1982).

The behavioral data presented here suggests that in many circumstances, humans can be shown to process speech stimuli differently from nonspeech stimuli and/or be more sensitive to between-category differences than to within category differs. Both sets of data would support the concept of a phonetic processor. Do humans always respond to speech as opposed to nonspeech sounds in different ways? Almost certainly not. There is no psychoacoustician on panel today, but he or she might argue that if you provide a sufficient amount of training to a human and wait for asymptotic performance after weeks of testing, you could almost certainly eliminate most, if not all categoricity from his/her processing of speech stimuli. Such data would reflect limits in the capability of the human auditory system—a measure of the abilities of mammalian auditory processing which may not differ in significant ways between humans and other mammals. However, knowledge of these absolute limitations provides only a very broad base on which to establish speech perception models. It would not demonstrate human proclivity in terms of how we naturally respond to speech stimuli as a function of normal development, speech communication context, and expectations. It is the proclivity of the human perceptual system to recognize speech as special and it is the proclivities of humans with regard to both the production and perception of speech sounds in which speech scientists and linguists should be most interested.

Notes

¹ For additional information, please contact first author at: Department of Speech and Hearing Science, The Ohio State University, 110 Pressey Hall, 1070 Carmack Rd., Columbus OH 43210-1002 (fox.2@osu.edu).

² These first three peaks may or may not correspond to those peaks identified as N100, P100 and N200 in many ERP studies examining cognitive responses to auditory stimuli.

³ Boatman, Lesser, and Gordon (1995) investigated auditory syllable discrimination, identification, and comprehension by direct cortical electrical interference. Under electrical interference at specific sites, auditory comprehension was selectively impaired, while syllable discrimination and identification remained relatively intact. This finding indicates that processes responsible for syllable discrimination and identification continue to operate despite impaired higher-level processes. Boatman, et al. suggest that general auditory processing cannot adequately explain why the subjects were able to selectively process acoustic-phonetic cues. Their data indicate that the neural resources subserving auditory speech processing are distributed differently with respect to their hierarchical status. Although auditory discrimination was localized to a single cortical site, syllable identification and auditory comprehension, two increasingly complex perceptual functions, were associated with additional sites, with comprehension localized at all sites tested. These neuroanatomic differences suggest that additional neural resources are required to support higher-level auditory speech functions.

References

- Aaltonen, O., Tuomainen, J., Laine, M. and Niemi, P. 1993. Cortical Differences in Tonal versus Vowel Processing as Revealed by an ERP Component Called Mismatch Negativity (MMN). *Brain and Language*, 44, 139-152.
- Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E., Lang, A.H. 1997. Perceptual Magnet Effect in the Light of Behavioral and Psychophysiological Data. *Journal of the Acoustical Society of America*, 2, 101-116.
- Bailey, P.J., & Herrmann, P. 1993. A Reexamination of Duplex Perception Evoked by Intensity Differences. *Perception & Psychophysics*, 54 (1), 20-32.
- Binder, J.R., Rao, S.M., Hameke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., and Haughton, V.M. 1994. Functional

- Magnetic Resonance Imaging of Human Auditory Cortex. *Annals of Neurology*, 35 (6), 662-672.
- Boatman, D., Lesser, R.P., and Gordon, B. 1995. Auditory Speech Processing In the Left Temporal Lobe: An Electrical Interference Study. *Brain & Language*, 51 (2), 269-290.
- Cohn, R. 1971. Differential Cerebral Processing of Noise and Verbal Stimuli. *Science*, 172, 599-601.
- Crain, S. 1991. Language Acquisition in the Absence of Experience. *Behavioral and Brain Sciences*, 17, 597-611.
- Creutzfeldt, O., Ojemann, G., and Lettich, E. 1989. Neuronal Activity in the Human Lateral Temporal Lobe. *Experimental Brain Research*, 77, 451-475.
- Cutting, J. and Rosner, B. 1974. Categories and Boundaries in Speech and Music. *Perception and Psychophysics*, 15, 591-600.
- Demonet, J.F., Chollet, R., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., and Frackwiak, R. 1992. The Anatomy of Phonological and Semantic Processing in Normal Subjects. *Brain*, 115, 1753-1768.
- Elman, J.L. Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. London: The MIT Press.
- ✓ Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- ✓ Fowler, C. A., and Rosenblum, L. D. 1992. Duplex Perception: A Comparison of Monosyllables and Slamming Doors. *Journal of Experimental Psychology: Human Perception and Performance*.
- Fox, R.A. 1982. Individual Variation in the Perception of Vowels: Implications for a Perception-Production Link. *Phonetica*, 39, 1-22.
- _____. 1983. Perceptual Structure of Monophthongs and Diphthongs in English. *Language and Speech*, 26, 21-60.
- _____. 1985. Auditory Contrast and Speaker Quality Variation in Vowel Perception. *Journal of the Acoustical Society of America*, 77, 1552-1559.
- _____. Flege, J.E. and Munro, M.J. 1995. The Perception of English and Spanish Vowels by Native English and Spanish Listeners: A Multidimensional Scaling Analysis. *Journal of the Acoustical Society of America*, 97, 2540-2551.
- Grossberg, S. 1976. Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors, *Biological Cybernetics*, 23, 121-134.
- Guenther, F.H. 1995. Speech Sound Acquisition, Coarticulation and Rate Effects in a Neural Network Model of Speech Production. *Psychological Review*, 102, 594-621.
- Guenther, F.H. and Gjaja, M.N. 1996. The Perceptual Magnet Effect as an Emergent Property of Neural Map Formation. *Journal of the Acoustical Society of America*, 100, 1111-1121.
- Hari, R. 1991. Activation of the Human Cortex by Speech Sounds. *Acta Otolaryngology*, Suppl. 491, 132-138.
- Gokcen, J. 1992. Neurological Evidence in Support of a Specialized Phonetic Processing Module. Unpublished Phd Dissertation, Department of Speech and Hearing Science, The Ohio State University, Columbus, OH.
- Hillyard, S. A. and Woods, D. L. 1979. Electrophysiological Analysis of Human Brain Function. In Gazzaniga (Ed.), *Handbook of Behavioral Neurology*, Vol. 2. New York: Plenum Press.
- Iverson, P. and Kuhl, P. 1995. Mapping the Perceptual Magnet Effect for Speech Using Signal Detection Theory and Multidimensional Scaling. *Journal of the Acoustical Society of America*, 97, 553-562.
- Jasper, H. H. 1958. The Ten-Twenty Electrode System of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371-375.
- Julesz, B. 1960. Binocular Depth Perception of Computer-Generated Patterns. *Bell System Technical Journal*, 39, 1125-1162.
- Julesz, B. 1971. *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.
- Kewley-Port, D. and Atal, B. 1989. Perceptual Differences Between Vowels Located in a Limited Phonetic Space. *Journal of the Acoustical Society of America*, 85, 1726-1740.

- Kimura, D. 1964. Left-Right Differences in the Perception of Melodies. *Quarterly Journal of Experimental Psychology*, 16, 335-358.
- Kimura, D. 1967. Functional Asymmetry of the Brain in Dichotic Listening. *Cortex*, 3, 163-178.
- Klatt, D. H. 1980. Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.
- Kleist, K. 1934. *Gehirmpathologie*. Joh-Ambros Barth, Leipzig.
- Kluender, K.R., Diehl, R.L., and Killeen, P.R. 1987. Japanese Quail Can Learn Phonetic Categories. *Science*, 237, 1195-1197.
- Kimura, D. 1964. Left-Right Differences in the Perception of Melodies. *Quarterly Journal of Experimental Psychology*, 16, 335-358.
- Kimura, D. 1967. Functional Asymmetry of the Brain in Dichotic Listening. *Cortex*, 3, 163-178.
- Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69.
- Kuhl, P. 1991. Human Adults and Human Infants Show a "Perceptual Magnet Effect" in the Perception of [I]: Identification and Change/No-Change Discrimination. *Journal of the Acoustical Society of America*, 97, 539-552.
- Kuhl, P. 1993. Innate Predispositions and the Effects of Experience in Speech Perception: the Native Language Magnet Effect. In B. De Boysson-Bardies, S. De Schonen, P. Jusczyk, P. Macneilage, and J. Morton (Eds), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, Kluwer Academic: Boston, Pp. 259-274.
- Kuhl, P. and Miller, P. 1975. Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants. *Science*, 190, 69-72.
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. 1992. Linguistics Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, 255, 606-608.
- Leslie, A.M. 1992. Pretense, Autism, and the "Theory of Mind" Module. *Current Directions in Psychological Science*, 1, 18-21.
- Liberman, A., Cooper, F., Shankweiler, D. and Studdert-Kennedy, M. 1967. Perception of the Speech Code. *Psychological Review*, 74, 431-461.
- Liberman, A., Harris, K., Hoffman, H. and Griffith, B. 1956. The Discrimination of Speech Sounds within and across Phoneme Boundaries. *Journal of Experimental Psychology*, 52, 127-137.
- Liberman, A. M. and Mattingly, I. G. 1985. The Motor Theory of Speech Perception Revised. *Cognition*, 21, 1-36.
- Liberman, A. M., Delattre, P. C., Cooper, F. C., and Gerstman, L. J. 1954. The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants. *Psychological Monographs*, 68, 1-13.
- Lightfoot, D. 1989. The Child Trigger Experience: Degree-0 Learnability. *Behavioral and Brain Sciences*, 12, 321-275.
- Luria, A.R. 1966. *Higher Cortical Functions in Man*. New York: Basic Books, Inc.
- Marler, P. 1970. Birdsong and Speech Development: Could There Be Parallels? *American Scientist*, 58, 669-673.
- Mann, V. A. and Liberman, A. M. 1983. Some Differences Between Phonetic and Auditory Modes of Perception. *Cognition*, 14, 211-235.
- Macmillan, N., Goldberg, R.F., and Braid, L.B. 1988. Resolution in Speech Sounds: Basic Sensitivity and Contact Memory on Vowel and Consonant Continua. *Journal of the Acoustical Society of America*, 84, 1262-1278.
- Mattingly, I. G., Liberman, A. M., Syrdal, A. K., and Halwes, T. 1971. Discrimination in Speech and Nonspeech Modes. *Cognitive Psychology*, 2, 131-157.
- Miller, J., Wier, C., Pastore, R., Kelly, W. and Dooling, R. 1976. Discrimination and Labelling of Noise Buzz Sequences with Varying Noise Buzz Lead Times: An Example of Categorical Perception. *Journal of the Acoustical Society of America*, 60, 410-417.
- Molfese, D. L. and Schmidt, A. 1983. An Auditory Evoked Potential Study of Consonant Perception in Different Vowel Environments. *Brain and Language*, 18, 57-70.

- Neville, H. J. 1974. Electrographic Correlates of Lateral Asymmetry in the Processing of Verbal and Nonverbal Auditory Stimuli. *Journal of Psycholinguistic Research*, 3, 151-163.
- Picton, T.W. and Stuss, D.T. 1984. Event-Related Potentials in the Study of Speech and Language: A Critical Review. In Caplan, Lecours, and Smith (Eds.) *Biological Perspectives on Language*, Cambridge, MA: MIT Press, Pp. 303-360.
- Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.
- Poggio, G. F. 1984. Processing of Stereoscopic Information in Primate Visual Cortex. In G. M. Edelman, W. E. Gall, and W. M. Cowan. (Eds.), *Dynamic Aspects of Neocortical Function*, New York: Wiley.
- Price, C.J., Wise, R.J.S., Warburton, E.A., Moore, C.J., Howard, D., Paterson, K., Frackowiak, R.S.J. and Friston, K.J. 1996. Hearing and Saying: the Functional Neuro-Anatomy of Auditory Word Processing. *Brain*, 119, 919-931.
- Ratliff, S. S. and Greenberg, H. J. 1972. The Averaged Encephalic Response to Linguistic and Nonlinguistic Auditory Stimuli. *Journal of Auditory Research*, XII (1), 14-25.
- Sams, M., Aulanko, R., Aaltonen, O. and Nääätänen, R. 1990. Mismatch Negativity to Changes in the Spatial Location of an Auditory Stimulus. *Electroencephalography and Neurophysiology*, 77, 436-444.
- Sharma, A., Kraus, N., Mcgee, T., Carrell, T. and Nicol, T. 1993. Acoustic Versus Phonetic Representation of Speech as Reflected by the Mismatch Negativity Event-Related Potential. *Electroencephalography and Clinical Neurophysiology*, 88, 64-71.
- Singh, S. and Woods, D.R. 1971. Perceptual Structure of 12 American English Vowels. *Journal of the Acoustical Society of America*, 49, 1861-1866.
- Sussman, J. and Lauckner-Morano, 1995. Further Tests of the "Perceptual Magnet Effect" in the Perception of [I]: Identification and Change/No-Change Discrimination. *Journal of the Acoustical Society of America*, 97, 539-552.
- Suga, N. 1984. The Extent to Which Binaural Information is Represented in the Auditory Cortex. In G.M. Edelman, W.E. Gall, & W.M. Cowans (Eds.), *Dynamic Aspects of Neocortical Function*. New York: Wiley.
- Takane, Y., Young, F., and Deleeuw, J. 1976. Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42, 7-67.
- Thorpe, W.H. 1958. The Learning of Song Patterns by Birds, With Special Reference to the Song of the Chaffinch, *Fringilla coelebs*. *Journal of the Acoustical Society of America*, 100, 535-570.
- Wall, L. G., Fox, R. A., Moenter, D., and Dalebout, S. D. 1991. Effect of Speech Distinctions and Age Differences on Auditory Event-Related Potentials. *Journal of the American Academy of Audiology*, 2,237-245.
- Whalen, D. H. and Liberman, A. M. 1987. Speech Perception Takes Precedence over Nonspeech Perception. *Science*, 237, 169-171.
- Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. 1992. Lateralization of Phonetic and Pitch Discrimination in Speech Processing. *Science*, 256, 846-849.