available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Electrophysiological evidence for early interaction between talker and linguistic information during speech perception

Natalya Kaganovich[a,*], *Alexander L. Francis*[a,b], *Robert D. Melara*[c]

[a] *Linguistics Program, Purdue University, West Lafayette, IN 47907-1353, USA*
[b] *Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN 47907, USA*
[c] *Department of Psychology, City College, City University of New York, NY 10031, USA*

ARTICLE INFO

Article history:
Accepted 14 July 2006
Available online 22 August 2006

Keywords:
Selective attention
Voice perception
Talker identification
Speech perception
Event-related potentials
P3

ABSTRACT

This study combined behavioral and electrophysiological measurements to investigate interactions during speech perception between native phonemes and talker's voice. In a Garner selective attention task, participants either classified each sound as one of two native vowels ([ɛ] and [æ]), ignoring the talker, or as one of two male talkers, ignoring the vowel. The dimension to be ignored was held constant in baseline tasks and changed randomly across trials in filtering tasks. Irrelevant variation in talker produced as much filtering interference (i.e., poorer performance in filtering relative to baseline) in classifying vowels as vice versa, suggesting that the two dimensions strongly interact. Event-related potentials (ERPs) were recorded to identify the processing origin of the interference: an early disruption in extracting dimension-specific information or a later disruption in selecting appropriate responses. Processing in the filtering task was characterized by a sustained negativity starting 100 ms after stimulus onset and peaking 200 ms later. The early onset of this negativity suggests that interference originates in the cognitive effort required by listeners to extract dimension-specific information, a process that precedes response selection. In agreement with these findings, our results revealed numerous dimension-specific effects, most prominently in the filtering tasks.

1. Introduction

Speech contains two separate sources of information: the talker's voice and the linguistic content of the utterance. Despite co-existing in a single auditory signal, these two sources support partially different cognitive functions. On the one hand, the human voice conveys information about the talker's identity, gender, age, health, and emotional state (Kreiman, 1997). Thus, talkers can be identified on the basis of non-speech vocalizations. On the other hand, the processing of linguistic content can proceed without the vocal qualities of the talker. Accordingly, the meaning of sine-wave speech can be understood, even though the spoken words have been deprived of their voice quality (Remez et al., 1981). The relative independence of linguistic and vocal information is revealed most clearly in recent research on the neural correlates of speech perception, which indicates that the two sources activate partially distinct brain areas (Belin et al., 2000, 2002; Belin and Zatorre, 2003; von Kriegstein et al., 2003; von Kriegstein and Giraud, 2004). One unanswered question, however, is when and how talker and language dimensions interact and combine in speech perception. The purpose of the

[*] *Corresponding author*. Fax: +1 765 494 0771.
E-mail address: kaganovi@purdue.edu (N. Kaganovich).

0006-8993/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2006.07.049

present study was to address this question using a combined behavioral and electrophysiological analysis.

1.1. Behavioral interactions between voice and language

Studies of list recall demonstrate that word memory is strongly affected by talker variation. Nygaard et al. (1995) contrasted recall of word lists spoken by the same or by different talkers with words presented at either short or long inter-stimulus intervals (ISIs). At short ISIs, multiple-talker lists were recalled significantly worse than single-talker lists, a pattern that reversed at longer ISIs. The authors concluded that the vocal qualities and the semantic content of words are encoded together in long-term memory. Transfer to long-term memory takes time, which is why recall is disrupted when multiple talkers follow one another at short ISIs.

Conversely, studies of auditory learning indicate that talker identification is strongly affected by linguistic content. Listeners who succeed at identifying a group of talkers on the basis of spoken sentences do not transfer this ability later to identifying the same talkers on the basis of words. Thus, talker-specific acoustic information depends on the type of linguistic information available to the listener (Nygaard and Pisoni, 1998). Although these results provide evidence that talker identity interacts with memory for spoken words, research on memory of spoken words cannot provide conclusive evidence regarding the role of talker identity in online speech perception.

1.2. Dimensional interactions in the Garner paradigm

One important body of behavioral research has investigated speech processes that precede memory and identification. Researchers in this area have adopted the Garner selective attention task (Garner and Felfoldy, 1970; Garner, 1974; Pomerantz et al., 1989) to study whether speech sounds and talkers' voices are processed together during perceptual encoding. If dimensions are processed *separately*, then individuals can classify stimuli along the target dimension as quickly when the irrelevant dimension changes randomly as when it stays constant or co-varies predictably with the target dimension. However, if the two dimensions are *integral* and processed jointly in perception, then random variation along the irrelevant dimension will delay target classification, an outcome called *filtering interference*. Moreover, co-variation between the target and irrelevant dimensions will speed classification, an outcome called *correlation gain*.

Mullennix and Pisoni (1990) used the Garner paradigm to study interactions between the dimensions of talker and consonant. They found that listeners were significantly slower at identifying the first consonant in a word when the talker changed randomly between males and females (filtering task) compared with when the talker remained constant (baseline task). Listeners also were delayed (but to a lesser extent) in identifying the talker's gender when the initial consonant changed randomly compared with baseline. These results suggested that the dimensions of talker and word are integral, perhaps influencing each other during the perception of the speech signal. Since the interference was asymmetrical, with variation along the talker dimension causing a larger response delay than variation along the consonant dimension, the authors suggested that the locus of interaction differed between talker and consonant, with talker information being available relatively earlier in processing.

1.3. Processing origins of filtering interference

Reports of filtering interference between talker and word are open to several different interpretations. One might consider filtering interference as arising early in perception, in the inability of perceivers to decompose holistically perceived stimuli into their dimensional constituents (Lockhead, 1972). On this view, correlation gain (in the correlated task) would accompany filtering interference (in the filtering task), because holistic entities differing on two dimensions are relatively distant in Euclidean perceptual space (Lockhead, 1966).

Alternatively, the locus of interaction may originate in working memory. The dimensions of talker and word are themselves multidimensional, defined by a common set of acoustical parameters, including formant frequency (Kreiman, 1997; Belin et al., 2004) and voice onset time (Swartz, 1992; Allen and Miller, 2004). Thus, the same set of formant frequencies can be heard as signaling either a specific voice or a specific speech sound, contributing to the difficulty of separating talker- and language-related information. Consequently, filtering interference may be due to the cognitive effort required to create separate categorical representations of dimensions and maintain these representations in working memory while performing the task. On this view, correlation gain may not accompany filtering interference because stimuli from a correlated set may be coded similarly in working memory as stimuli from a baseline set.

The modal model of verbal working memory postulates the existence of separate components for phonological, semantic, and syntactic information (Baddeley, 1986, 1990; Martin and Romani, 1994). Accordingly, interactions between talkers and words may occur at different levels of linguistic analysis (Melara and Marks, 1990). Eisner and McQueen (2005) found evidence for phonemic integrality in listeners who monitored speech at a talker-specific phonemic level. They argued that listeners often are able to use talker identity to disambiguate phonemic cues. Green et al. (1997) used the Garner paradigm to investigate the relationship between talkers' voice, speaking rate, and consonant voicing. In contrast to Mullennix and Pisoni (1990), these authors obtained evidence of separability between the talker and linguistic dimensions. One difference between these two studies is that Mullennix and Pisoni used as stimuli short meaningful words that formed minimal pairs whereas Green, Tomiak and Kuhl used meaningless syllables. The differences in stimuli and outcomes suggest that the interactions between talker and language dimensions may be limited to relatively late in processing, only after the semantic content of a word is extracted.

Finally, interactions between talker and word may arise from disruptions occurring during response selection. Performance in the Garner paradigm is best if the observer can reduce the four stimuli present in the filtering task (e.g., if classifying sounds based on pitch and loudness: high/loud, low/loud, high/soft, low/soft) to two values of the classification dimension (e.g., high pitch vs. low pitch), thereby converting a many-to-one mapping of stimuli-to-responses (e.g., high/loud or high/soft to left key, low/loud or low/soft to right key) to a one-to-one mapping (high to left key, low to right key). Since the baseline task (with two stimuli) is always a one-to-one mapping, an observer who can reduce the filtering task from four choices to two choices has achieved separability. This goal may be especially difficult to reach when stimulus features are physically interrelated, as in the case of talker and word, requiring observers during filtering to laboriously map four physically overlapping stimuli to two distinct responses. Such involved response selection processes would slow classification in the filtering task relative to baseline, leading to filtering interference.

Extant behavioral research has been equivocal on the locus of interaction between talker and linguistic dimensions. As discussed, the interaction may occur early in perception, during the formation of sound representations and their maintenance in working memory, or during response selection. Moreover, the locus of interaction may have multiple origins or, as Mullennix and Pisoni's (1990) finding of asymmetrical interference shows, may change depending on the dimension of classification (i.e., variation in talker may cause a larger response delay than variation in word). Behavioral measures alone are unable to distinguish among these possibilities.

1.4. *Using electrophysiological analysis to identify the locus of interaction*

The primary purpose of the present study was to use an electrophysiological analysis to pinpoint more accurately the time course of processing and interaction between talker and linguistic dimensions. Event-related potentials (ERPs) provide a means of isolating stimulus evaluation and response selection processes, and thereby aid in determining the cognitive processes responsible for interference and response delay. Kutas et al. (1977) suggested that the latency of the P3[1] component in the ERP waveform indexes the time necessary for stimulus evaluation and is largely independent of the response selection processes. McCarthy and Donchin (1981) directly tested this hypothesis by manipulating the difficulty of identifying and responding to a stimulus as response latency and the latency of the P3 component were measured. The results were consistent with the view that P3 latency is affected by stimulus evaluation and categorization but not by the difficulty of response selection. More recently, Klostermann et al. (2006) showed that the cortical P3 response is preceded by a subcortical component elicited by a stimulus' status as target (vs. non-target), whether or not observers actually respond to that stimulus. Such findings have been confirmed by other research (Magliero et al., 1984; Dien et al., 2004; Calhoun et al., 2006). Although the nature of the ERP components underlying the P3 and the type of tasks that may alter its sensitivity to perceptual or post-perceptual processes are still a topic of debate (Verleger, 1997; Leuthold and Sommer, 1998), the assumption that the latency of the P3 primarily indexes perceptual processes has been used successfully to study a variety of issues, including attention (Duncan-Johnson, 1981; Lew et al., 1997), clinical testing (Polich and Herbst, 2000), and aging (Bashore and Ridderinkhof, 2002). Because the P3 occurs after a stimulus has been categorized, it can serve as an extremely useful tool in determining the locus of interaction between talker- and language-related information during speech perception.

1.5. *The present study*

This study combined the Garner selective attention task with simultaneous ERP recording in order to determine the locus and nature of interaction between a talker's voice and phoneme. We measured ERPs as listeners classified talkers or isolated phonemes in the Garner paradigm. The primary goal was to ascertain whether dimensional interaction occurred early, during stimulus evaluation and categorization, or late, during response selection (Smith and Kemler, 1978; Smith and Kilroy, 1979).

Participants performed vowel classification (two American English vowels) and talker classification (two male talkers) in baseline tasks (irrelevant dimension held constant), filtering tasks (irrelevant dimension changed randomly), and correlated tasks (relevant and irrelevant dimensions co-varied). Reaction times, accuracies, and ERP recordings were collected during each task. Since a Garner-style selective attention task requires listeners to categorize each heard sound into one of two groups, the paradigm is suitable for eliciting the P3 component, which typically is observed in any task requiring a binary decision (Kutas and Van Petten, 1994).

By comparing ERP recordings during baseline and filtering tasks, and by using the latency of the P3 component to mark the end of perceptual processing, we were able to probe whether a change in the irrelevant dimension affected the nature of sound encoding, the ease of response selection, or both. If the interaction between the dimensions of talker and vowel occurs during early sound encoding, then baseline and filtering ERP recordings should differ significantly from each other preceding the onset of P3. If, however, the dimensions interact primarily during response selection or other post-perceptual processes, the difference in baseline and filtering ERP recordings should occur after the P3 component.

Lew and colleagues (1997), who were among the first to use electrophysiological measurements to investigate the effect of randomly changing a linguistic dimension on the auditory processing of words, concluded that a combination of early and late processes contributed jointly to filtering interference. However, their study did not address the effect of talker change on the encoding of linguistic information and, similar to earlier behavioral studies, defined the talker dimension by a gender difference. By focusing on the two-way interaction between linguistic and paralinguistic aspects of the speech signal, and by defining talkers within gender, the current study significantly extends our understanding of the cognitive processes underlying speech perception.

[1] Different conventions exist for naming ERP waveform components. A given component can be labeled by the exact peak latency identified in a specific study (e.g., P370), by a conventional latency established in the literature (e.g., P300), or by its order in the sequence of components (e.g., P3). For simplicity and consistency, we adopted the sequence nomenclature throughout this article.

2. Results

2.1. *Behavioral findings*

Repeated-measures ANOVAs were conducted separately on reaction times and accuracies in order to compare participants' performance across the three tasks (baseline, filtering, and correlated) and two dimensions (talker and vowel). The analysis of reaction times showed a significant effect of task, $F(2,42)=187.9$, $p<0.001$, but no effect of dimension, $F(1, 21)=3.1$, *ns*, and no task by dimension interaction, $F(2,42)=0.37$, *ns*. A Tukey post-hoc analysis revealed that in both dimensions, the filtering task produced significantly longer reaction times than the baseline task, with an increase from baseline to filtering of 152 ms in the talker dimension, $p<0.001$, and 166 ms in the vowel dimension, $p<0.001$. Additionally, neither dimension showed a significant difference between the baseline and correlated tasks: talker, 17 ms; vowel, 13 ms. Performance on talker and vowel baseline tasks did not differ significantly from each other (Fig. 1).

The analysis of accuracy showed a significant effect of task, $F(2, 42)=28.33$, $p<0.001$, but no effect of dimension, $F(1, 21)=3.83$, *ns*. A Tukey post-hoc analysis showed that participants' performance was significantly less accurate in the filtering task compared to both the baseline and the correlated tasks, which did not significantly differ from each other (Fig. 2). There was also a significant dimension by task interaction, $F(2, 42)=4.12$, $p<0.05$, with talker classification being less accurate than vowel classification in the filtering task. Other tasks did not show significant differences between the two dimensions.

2.2. *ERP findings*

2.2.1. *Garner effects*

Fig. 3 depicts ERP waveforms from the baseline and filtering tasks, with data from both dimensions combined.

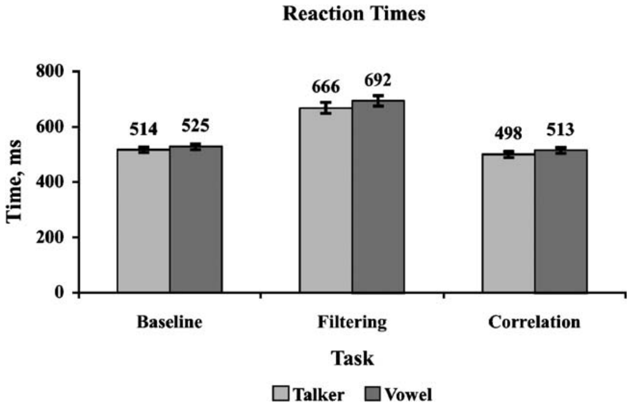

Fig. 1 – Comparison of reaction times in baseline, filtering, and correlated tasks across talker and vowel dimensions. Numbers on top of graphs show average reaction times for performing a given task ($n=22$). Error bars reflect standard error.

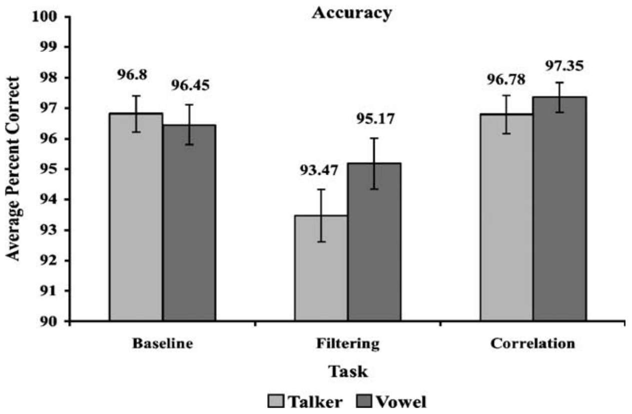

Fig. 2 – Comparison of accuracies in baseline, filtering, and correlated tasks across talker and vowel dimensions. Numbers on top of graphs show average percent correct for identification of stimuli in a given task ($n=22$). Error bars reflect standard error.

The first significant bioelectric difference between the baseline and filtering tasks occurred in the N1 component over parietal and temporal electrodes, $F(2, 42)=7.0$, $p<0.01$, which was relatively greater in amplitude in the filtering task (Fig. 4).

Garner effects were also observed in the components following the N1 component. The peak amplitude of the subsequent positive components, P2 and P3, was significantly reduced in the filtering task compared with baseline [P2, $F(2, 42)=5.93$, $p<0.01$ over parietal and temporal electrodes; P3, $F(2, 42)=8.36$, $p<0.001$], whereas the amplitude of the N2 component was significantly enhanced [$F(2, 42)=13.8$, $p<0.001$]. Thus, enhanced voltage negativity characterized ERP waveforms in the filtering task across the N1, P2, N2, and P3 components. Lastly, the terminal leg of the P3 component, occurring 500–700 ms after stimulus onset, was significantly more positive in the filtering task, $F(2, 42)=6.13$, $p<0.01$.

2.2.2. *Dimension-specific effects*

We found a number of significant dimension-specific effects in the electrophysiological results that were present in both the peak amplitude and latency of three positive ERP components—P1, P2 and P3. The earliest significant difference between talker and vowel dimensions occurred in the latency of the P1 component: In the filtering task, P1 peaked significantly earlier in the talker dimension than in the vowel dimension, $F(1, 21)=4.91$, $p<0.05$ (Fig. 5).

The peak amplitude of the P2 component was significantly reduced in the talker dimension compared with the vowel dimension over parietal and temporal electrodes, $F(1, 21)=8.69$, $p<0.01$. Additionally, the peak of the P2 component occurred significantly later in the right-hemisphere compared with the left-hemisphere in the vowel dimension over central, parietal, and temporal sites as revealed in the dimension by hemisphere interaction, $F(1, 21)=16.25$, $p<0.001$ (Fig. 6).

Lastly, in the filtering task, the P3 component was significantly larger in the talker dimension over frontal, central, and parietal sites, $F(1, 21)=4.63$, $p<0.05$ (Fig. 7), and

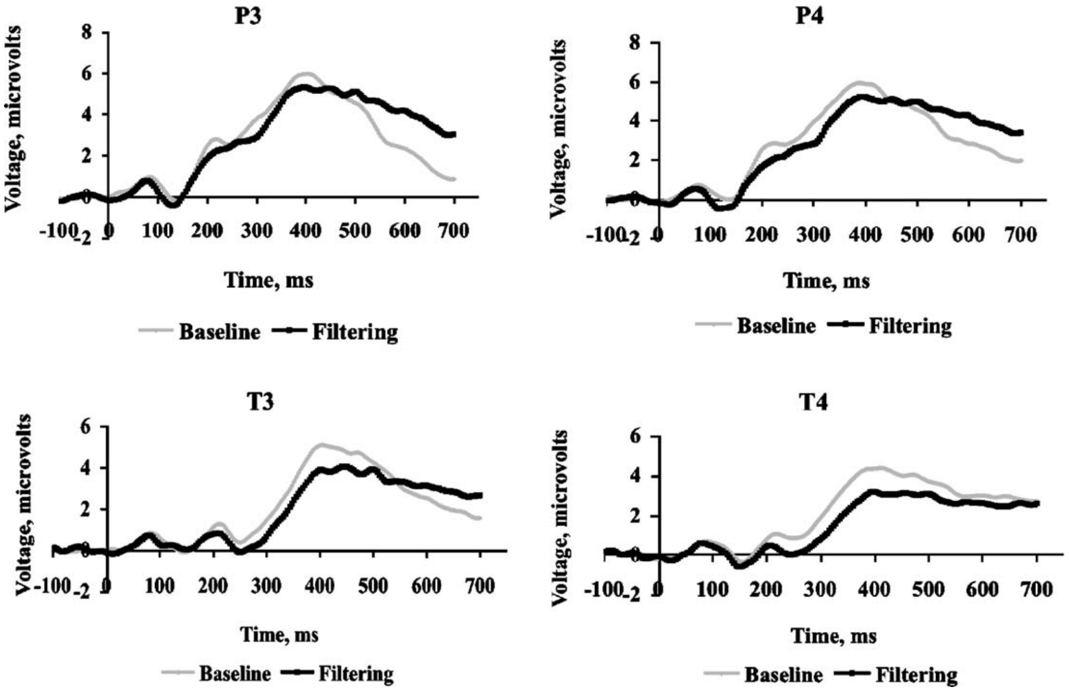

Fig. 3 – Comparison of baseline and filtering tasks across both dimensions. Waveforms represent group data averaged across all 22 participants. Only parietal and temporal sites are shown. All waveform graphs show time in milliseconds along the *x*-axis and voltage in microvolts along the *y* axis. Medial electrodes are not shown.

occurred significantly later compared with the vowel dimension across all sites, $F(1, 21)=4.82$, $p<0.05$ (Fig. 8).

3. Discussion

3.1. Behavioral findings

Our behavioral measurements were designed to test whether the dimensions of talker and vowel are processed conjointly or separately during perception. For both dimensions, we found significantly poorer performance in both reaction time and accuracy in the filtering task compared with the baseline task (Figs. 1 and 2) indicating *filtering interference*. Within the framework of the Garner paradigm, this result suggests that the dimensions of talker and vowel are integral. The filtering interference was fully symmetrical, with random variation in vowel producing as much delay in the classification of talkers as random variation in talker did in the classification of vowels. Reaction times did not differ significantly between the two dimensions in any of the tasks, indicating that identifying talkers and identifying vowels were comparable in difficulty.

Neither dimension showed a *correlation gain* in either accuracy or reaction time: When the dimensions varied in a correlated manner (i.e., the correlated task), participants did not capitalize on the simultaneous change in vowel and talker to improve their performance relative to baseline. The absence of correlation gain cannot be explained by a floor effect in participants' responses because reaction times in the baseline tasks exceeded 500 ms. Although slightly unusual, this finding replicates the results that Mullennix and Pisoni (1990) obtained in their 2×2 condition. A possible explanation lies in the complexity of dimensions used in the two studies as

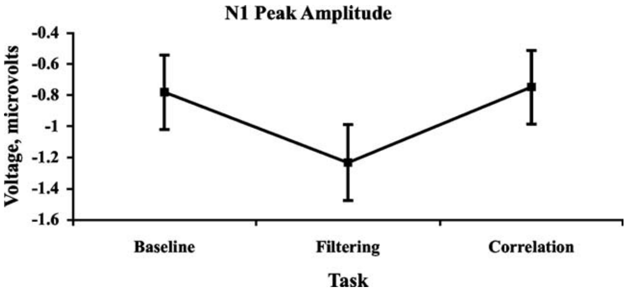

Fig. 4 – The effect of task on the amplitude of N1 component across both dimensions over parietal and temporal sites. Data points represent group data averaged across all 22 participants. Error bars reflect standard error.

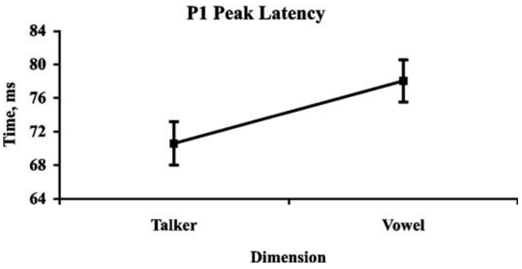

Fig. 5 – P1 latency difference between talker and vowel dimensions in the filtering task. Data points represent group data averaged across all 22 participants. Error bars reflect standard error.

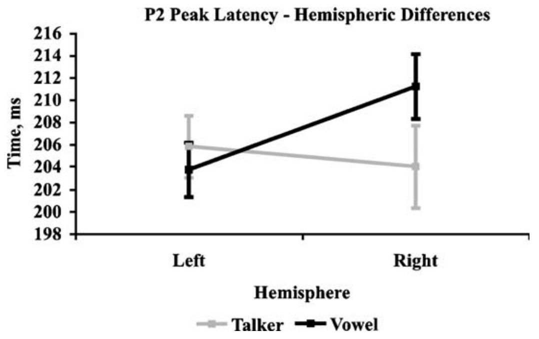

Fig. 6 – Hemispheric difference in P2 peak latency between dimensions over central, parietal, and temporal sites. Data points represent group data averaged across all 22 participants. Error bars reflect standard errors.

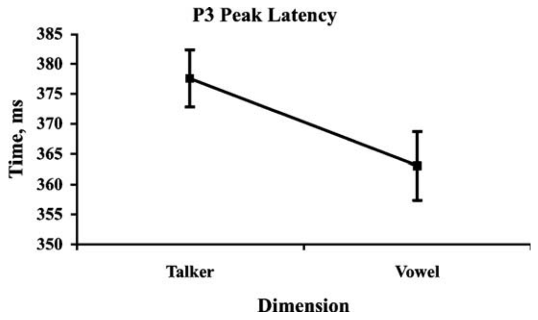

Fig. 8 – P3 latency difference between talker and vowel dimensions in the filtering task. Data points represent group data averaged across all 22 participants.

compared with those typically used in Garner tasks. A correlation gain is usually attributed to the enhanced psychological distance between stimuli that differ along two dimensions. However, when the two dimensions are themselves multidimensional and share a number of acoustic properties that are categorized differently depending on the task, as the talker and vowel dimensions are in the present study, then a simple summation of differences between dimensions may not yield greater inter-stimulus distance. This issue requires further study.

To the best of our knowledge, this is the first study to show a processing dependence in talker- and language-related information at the phonemic level. Our results are in agreement with the findings of Eisner and McQueen (2005) who demonstrated that listeners are able to monitor speech for talker-specific information at the level of individual phonemes. Additionally, we were able to demonstrate the interaction between talkers' voice and phonemes by using within-gender talker variation. This result indicates that a talker change interferes with classification along the linguistic dimension even in the absence of the large acoustic and categorical difference between male and female voices used in earlier studies (Mullennix and Pisoni, 1990; Green et al., 1997).

Mullennix and Pisoni (1990) found that talker change was more disruptive to the identification of the voicing of the first consonant in a word than vice versa. They concluded that

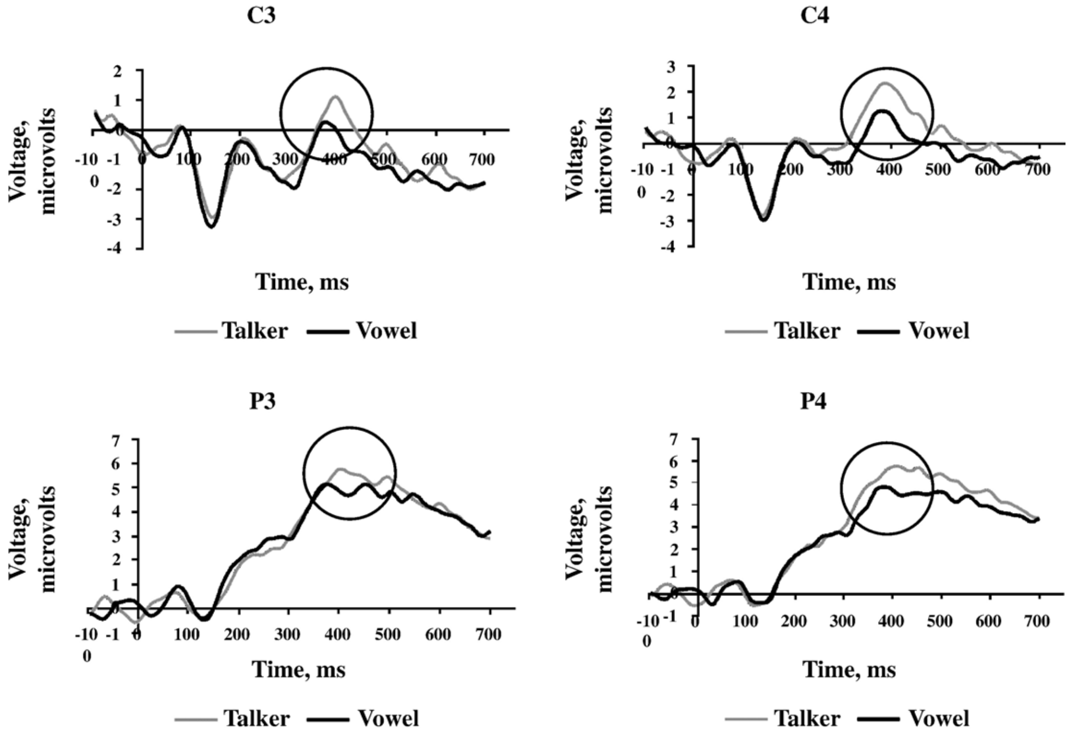

Fig. 7 – P3 peak amplitude differences between talker and vowel dimensions in the filtering task. Waveforms represent group data averaged across all 22 participants. Only central and parietal sites are shown. All waveform graphs show time in milliseconds along the *x*-axis and voltage in microvolts along the *y* axis. Location of the P3 peak is identified with a circle. Medial electrodes are not shown.

phonemic processing depends inherently on talker identity. By contrast, we found a symmetrical dependency between talkers and vowels. At least three factors might have contributed to the differences in findings. First, in Mullennix and Pisoni's study the talker dimension varied between male and female voices, usually a very pronounced and acoustically salient difference. A decision about voice gender may be possible within a few tens of milliseconds following the burst release of the consonant (Swartz, 1992) or, if the decision is based on voice pitch, within a few fundamental frequency cycles (a few tens of milliseconds more) into the vowel following the first consonant (Robinson and Patterson, 1995). If gender identification was perceptually easier than consonant identification, then irrelevant variation in talker would tend to create greater interference in consonant identification than vice versa. Conversely, Mullennix and Pisoni's use of short meaningful words may have slowed consonant identification: Listeners may have been able to make a consonant decision only after most of the word, including the vowel, was heard (Wood and Day, 1975; Miller and Dexter, 1988). If so, then the results of Mullennix and Pisoni may reflect the relationship between word and talker more than it does the relationship between phoneme and talker, as in the current study.

Lastly, the difference in findings between studies may have been due to a qualitative difference in the processing of consonants and vowels, with different acoustic cues for talker classification available to the two groups of participants. Different kinds of acoustic cues are used for consonant versus vowel identification (Raphael, 2005; Stevens, 2005). These cues unfold differently over time, and processing of the two kinds of sounds may even rely on distinct neural mechanisms (Caramazza et al., 2000). The nature of acoustic cues used for talker classification or the relative weight of such cues, as well as their dependence on a particular listener and a particular situation, remains an open question (Kreiman, 1997; Lavner et al., 2000; Kreiman et al., 2005).

3.2. *ERP findings*

The purpose of the electrophysiological analysis was to determine the temporal locus of interaction between the dimensions of talker and vowel. We used the P3 component as an index of the end of the evaluation stage of sound processing. We compared ERP waveforms in the baseline and filtering tasks to identify the processing origins of filtering interference. We reasoned that if the interaction between talker and vowel occurs during early sound encoding, then the differences between ERP waveforms in the two tasks would be found in early ERP components, prior to P3. Conversely, if the delay in classification during filtering results from disruption at the response selection stage of the task, then the differences between tasks would occur after the P3 component.

3.2.1. *Task-specific effects*

In concert with the behavioral findings, the ERP differences between tasks were symmetrical across the two dimensions. In both talker classification and vowel classification we found relatively greater voltage negativity in the filtering task across the N1, P2, N2, and P3 components. The negative enhancement was especially prominent in the N2 and P3 components over central electrode locations (Fig. 7). Because these task-specific effects were not associated with an individual ERP component, we believe that they reflect sustained attentional processes operating during performance of the filtering task. This sustained negativity may indicate the allocation of resources required for attentional effort, perhaps analogous to the processing negativity observed in dichotic listening tasks (Näätänen, 1990).

The early onset of the sustained negativity suggests that the attentional processes active in the filtering task preceded the selection of response codes (Ritter et al., 1979; Donchin and Coles, 1988; Picton, 1992; Oostenveld et al., 2001). It is likely that sustained attention was needed during filtering to extract the dimensional information required for classification decisions. For example, in vowel classification, observers performing the filtering task decided which of two categories each of the four sounds fell into: [ɛ] or [æ]. Yet some of the acoustic properties needed to perform vowel classification were identical to some of those distinguishing talkers. Thus cognitive effort was required to extract the acoustic information of the signal appropriate to the categories of the task-relevant dimension. We believe that the sustained negativity observed in the filtering task is an ERP signature of this attentional effort. Of course, effort was ultimately in the service of response selection: once the perceiver identified the dimensional category of a presented sound, the response associated with that category could be selected. However, the relatively early latency of the ERP components affected by our task manipulation suggests that the processing origins of interference effects occurred well before response selection.

The differences between the baseline and filtering tasks were greatest across the N2 and P3 ERP components, suggesting that attentional effort was directed at representations held in working memory (Novak et al., 1990; Pritchard et al., 1991; Polich and Kok, 1995; Polich and Herbst, 2000). It is plausible that category representations relevant to the task were maintained in working memory; the goal of sustained attention was thus to match a presented sound to its dimensional representation. This goal was more difficult to achieve in the filtering task than in baseline because double the number of sounds (four) must be matched to the same number of categories (two), thereby placing greater attentional demands on working memory. In the current study, we found that the terminal leg of the P3 component (500–700 ms after stimulus onset) was larger (more positive) over parietal sites in the filtering task (Fig. 3). One interpretation is that working memory processes lasted relatively longer in the filtering task because of the greater uncertainty in matching a sound to its dimensional representation[2]. Evidence of filtering interference in the behavioral measures suggests that our participants experienced a substantial processing cost when performing the filtering task, despite the presumed allocation of attentional resources. It has been claimed that Garner interference arises when holistic perceptual representations of the four stimuli in the filtering task are mapped without dimensional analysis onto a smaller number of response categories (Lockhead, 1972, 1979). On this view, one can

[2] Alternatively, latency jitter of the P3 response may have been relatively greater in the filtering task.

interpret the sustained ERP negativity observed in the filtering task reflecting the laborious re-mapping of unanalyzed wholes, rather than the effort expended to garner dimensional information. This holistic interpretation is implausible, however, because it suggests that perceivers are unable to detect the difference between a change in talkers and a change in vowels (Melara et al., 1993). Self-reports of our participants indicate that classifications were, in fact, strongly dimensional, with participants actively distinguishing the two talkers in talker classification, or the two vowels in vowel classification. Moreover, as we discuss next, several ERP results from this study confirm that dimension-specific characteristics of the signal were successfully derived during the course of stimulus classification. We contend that the extraction of these dimension-specific qualities was the product of early and sustained attentional effort, which ultimately came at the cost of classification speed and accuracy.

3.2.2. *Dimension-specific effects*

Our analysis of task effects revealed a striking symmetry in processing between the talker dimension and the vowel dimension. For both the behavioral measures and the ERP measures, differences between the baseline and filtering tasks were similar across the two dimensions, whereas differences between the baseline and correlated tasks were minimal between the two dimensions. However, ERP analyses performed within each task revealed several important dimension-specific effects. These effects were most prominent in the filtering task.

The earliest difference between the two dimensions was observed in the latency of the P1 component of the filtering task, with P1 occurring earlier during talker classification than during vowel classification (Fig. 5). A number of studies that have focused on the interaction between talker's voice and linguistic information suggest relatively earlier processing of the human voice in the auditory stream, with linguistic processing being contingent upon it (Mullennix and Pisoni, 1990; Eisner and McQueen, 2005). However, to the best of our knowledge, ours is the first study to provide electrophysiological evidence to this hypothesis. The P1 latency difference between dimensions implies that talker information is registered by the auditory system earlier than vowel information. An earlier detection of the voice dimension may be due to the intrinsic significance of voice information. The human voice contains important information about a person's identity, gender, and age. It provides unique insight into the emotional state of a human being, information fundamental to human social interactions. The importance of vocalizations for everyday social functions and survival mechanisms has been well documented for both non-human primate and non-primate species (Hauser, 1997; Ghazanfar and Hauser, 1999, 2001; Ghazanfar et al., 2001). Therefore, evolutionary selection pressures may have led to the early detection and processing of voices in auditory input.

Some support for this suggestion comes from a growing number of studies in humans and monkeys on the existence of brain areas that are specialized for face processing (e.g., Puce et al., 1995; Kanwisher et al., 1997; Tsao et al., 2006). An early ERP component, the N170, and its magnetic counterpart, the M170, have been shown to be sensitive to the presence of faces in visual stimuli (Bentin et al., 1996; Heisz et al., 2006; Xu et al., 2005). Since voice is often referred to as the "auditory face" (e.g., Belin et al., 2002), the existence of an early ERP component sensitive to vocal information would not be surprising. However, because voice was present in both dimensions in the current study and only the direction of attention differentiated between them, the exact nature of the P1 latency difference cannot be established at this time. Further studies would need to test whether this component reflects voice-specific processing in the brain.

The two dimensions also differed in the amplitude of the P2 component, approximately 200 ms after stimulus onset. The P2 effect was evident in each task, but was restricted to parietal and temporal electrodes. Vowel classification led to a larger P2 peak than talker classification. The P2 component has been associated with the maintenance and indexing of representations in working memory. Hence, the greater salience of vowel information revealed by the P2 effect might indicate more efficient retrieval of long-term memory traces for native language phonemes into working memory. As the two voices used in this experiment were unfamiliar to the participants, and thus were not represented in their long-term memory, no retrieval of voice-specific identity was possible during talker classification tasks.

We also found that, in all three tasks, the peak latency of the P2 component occurred earlier in left hemisphere sites than in right hemisphere sites during vowel classification (Fig. 6). This latency effect suggests a functional asymmetry between cerebral hemispheres, with the left hemisphere processing target sounds more quickly than the right whenever the task is to extract linguistic information. Although hemispheric asymmetry in relationship to language processing is a topic of ongoing debate (Shtyrov et al., 2005), a number of studies have demonstrated the presence of native phoneme memory traces in left auditory cortex (Tervaniemi et al., 2000; Shestakova et al., 2003; Makela et al., 2003; Liebenthal et al., 2005). Our findings are in agreement with this research.

Lastly, the difference between the two dimensions was also manifested in the peak amplitude and latency of the P3 component. The P3 component was significantly smaller in amplitude and peaked significantly earlier in the vowel-filtering task than in the talker-filtering task over frontal, central and parietal sites (Figs. 7 and 8). One interpretation of the difference in P3 amplitude and latency between dimensions is that a random change along the talker dimension was more detrimental to vowel classification than vice versa, suggesting that when talkers change, vowel representations becomes less certain. It is unclear, however, why a difference in the saliency of representations between the two dimensions did not manifest itself in the behavioral results.

An alternative interpretation is that the larger P3 peak generated during the talker-filtering task reflects the Voice-Sensitive Response (VSR), as proposed by Levy et al. (2003). Peaking at approximately 320 ms after stimulus onset, this positive component is elicited in response to human non-linguistic sounds (voices singing different musical notes), but not to the timbre of different instruments playing the same notes. Levy et al. claim that the VSR gauges the unique ability of the human voice to capture attention, possibly indexing speaker identification processes. Similarly, our P3 effect may also indicate processing related to talker identity. The scalp distribution of our response matches the fronto-central peak of activity reported by Levy et al. (2003). Although the P3 component elicited in the current study peaked on average at 380 ms post-stimulus onset, which is later than suggested by Levy et al. (2003), the delay may be due to the greater complexity of our stimuli.

3.3. Conclusions

We investigated the nature of interaction between speech sounds (vowels) and talker's voice in a Garner selective attention task. Behavioral and electrophysiological findings revealed a symmetrical pattern of dependency between dimensions, with a random change in talker providing as much interference in the classification of vowels as vice versa. By comparing electrophysiological recordings during baseline and filtering tasks, and using the P3 component as the landmark of the end of perceptual processing, we have found that the filtering task was characterized by greater voltage negativity 100–300 ms after stimulus onset compared with baseline. This sustained negativity started with the N1 component over parietal and temporal sites and was most prominent over the N2 and P3 components of the waveform. The early onset of this negativity suggested that attentional effort in the filtering task started well before response selection and likely reflected the effort needed to extract dimension-specific information in order to correctly perform the task. In agreement with this interpretation, we found a number of dimension-specific ERP effects, most prominently in the filtering task. Taken together, our findings indicate that whereas dimensions of speech and talker are strongly interconnected during perception, healthy normal individuals are able to extract dimension-specific information from multi-dimensional acoustic signal with high accuracy. The results are consistent with a contingent parallel model of language processing (Mullennix and Pisoni, 1990; Knosche et al., 2002), in which the processing of phonemes depends upon the prior processing of voice, even when vowels and talkers are matched in perceptual discriminability.

4. Experimental procedure

4.1. Participants

Twenty-four right-handed undergraduate students (9 female, 15 male) enrolled in an introductory psychology class participated for course credit. All gave their written consent to participate in the experiment, which was approved by the Institutional Review Board of Purdue University. All participants were administered a brief audiological exam to ensure normal hearing. Each participant completed the Edinburgh Handedness Inventory to assess handedness (Oldfield, 1971). The data of two participants (1 female, 1 male) were excluded from the analysis. One participant lacked the exogenous N1 component in her ERP waveforms possibly due to a subclinical auditory problem. Noise artifacts contaminated the ERP recordings of another participant.

4.2. Stimuli

Four sounds were used as stimuli in this study: The vowels [ɛ] and [æ] produced by two male talkers of the Midwestern dialect of American English. Talkers produced each vowel in isolation. All sounds were recorded using a digital audio tape-recorder (Sony TCD-D8) and hypercardioid microphone (Audio-Technica D11000HE) in a single-walled sound-isolating booth (IAC, Model #403A). Steady-state portions of vowels were extracted from recordings, peak amplitude normalized and adjusted to 180 ms in duration using Praat 4.3. To ensure gradual attack and decay, each sound's amplitude was artificially ramped from 0 to maximum over the first 30 ms, and from maximum to 0 over the last 30 ms. Five milliseconds of silence were added to the start and 10 ms of silence were added to the end of each sound. All stimuli were given a slightly falling fundamental frequency (f0) contour by using the pitch tier manipulation method in Praat 4.3. The amount of f0 fall between the first and the last pitch points for each sound was 15 Hz but the talker's average f0 was maintained for each sound. To identify appropriate stimuli for this experiment, a number of different vowels were first recorded from several different talkers and tested on native English-speaking volunteers (members of the lab) in a 2-AFC identification paradigm. The final choice of stimuli was made based on the results of these tests, such that the discriminability of the selected vowels and talkers was as close to equal as possible (Fig. 9, Panel A).

Although the reaction times shown in Fig. 9 are not identical, the difference between talker and vowel classification is negligible when opposite sides of the square are averaged: talker dimension: 392 ms (SE = 20 ms) versus vowel dimensions: 402 ms (SE = 12 ms), Fig. 9, Panel B. Because this study was concerned with identifying vowels and talkers as categories rather than as specific exemplars, we averaged data across both tokens of each category during data analysis.

4.3. Task

In each task, participants were familiarized with the stimulus set prior to identifying the sounds, presented one at a time, as either one of the two vowels (vowel classification) or one of the two talkers (talker classification). Participants performed five tasks along each dimension of classification: two baseline tasks, one filtering tasks, and two correlated tasks (Table 1).

In each baseline task, the stimulus set contained two sounds differing along the relevant dimension, with the irrelevant dimension kept constant. For example, in the talker baseline participants were asked to identify the talker of a single vowel (e.g., [ɛ]). In each filtering task, the stimulus set contained four sounds (i.e., [ɛ] and [æ] spoken by Talker 1 or Talker 2) sorted into two groups depending on the dimension of classification (Table 1). Thus, values along the irrelevant dimension changed randomly from trial to trial in filtering. For example, in talker filtering, participants classified the sounds as Talker 1 or Talker 2 regardless of the vowel spoken. In correlated tasks, the stimulus set contained two sounds, which differed from each other along both dimensions. Thus, the irrelevant dimension changed predictably with the

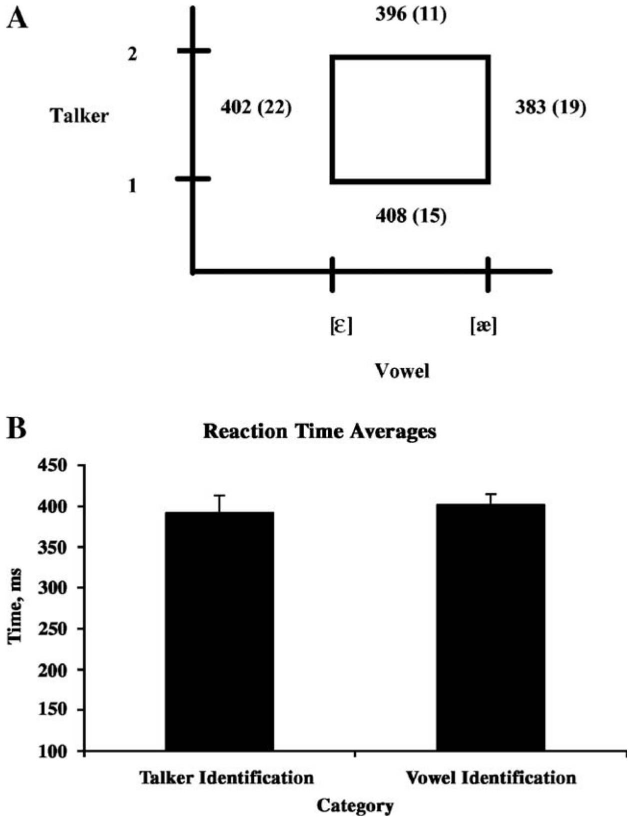

Fig. 9 – (A) Perceptual distances between four stimulus sounds. Numbers by the square sides are reaction times for identifying either two vowels or two talkers averaged across six listeners. Numbers in parentheses are standard errors of the mean. For example, in the vowel classification task, it took participants on average 408 ms to identify vowels [ɛ] and [æ] spoken by talker 1, and it took them on average 402 ms to identify talkers 1 and 2 based on the vowel [ɛ], etc. (B) Reaction time difference between talker and vowel identification after the opposite sides of the perceptual square were averaged together. Error bars represent standard error.

dimension of classification. Performance on the baseline and correlated tasks is reported as the average of two tasks along the dimension of classification.

Participants were seated in a comfortable chair in a sound-attenuating chamber (Industrial Acoustics Company, New York). Stimuli were presented to listeners binaurally over headphones (Nova-40) at a comfortable listening level. Participants were instructed to respond as quickly as possible by pressing one of two computer keys with the forefinger and the middle finger of their right hand. Each task consisted of 96 trials (48 for each of two sounds) preceded by 10 practice trials. Responses to practice trials were not recorded. Presentation of sounds was randomized within tasks; order of tasks and mapping of response keys was counterbalanced across participants. Each participant classified sounds along both dimensions, with half beginning with talker classification and half with vowel classification. Each experimental session lasted approximately 1.5 h, in addition to EEG preparation.

4.4. *ERP recording*

In each task, the electroencephalogram (EEG) was recorded from 13 scalp locations (Fz, F3, F4, Cz, C3, C4, Pz, P3, P4, T3, T4, T5, and T6 of the international 10–20 System) and the left mastoid (LM) with tin electrodes mounted in a stretch cap (Electro-Cap International). All scalp electrodes were referenced to an electrode positioned on the tip of the nose. Fpz served as the ground electrode. Impedance was maintained below 1.5 kΩ across all sites. Blinks and other eye movements were monitored by electrooculogram (EOG) from two electrode montages, one on the infra- and supra-orbital ridges of the left eye (VEOG), the other on the outer canthi of each eye (HEOG).

4.5. *Data analysis*

4.5.1. *Behavioral data*

Within each dimension, reaction times for all correct identifications of sounds were averaged across all participants for a given task. Additionally, the percentage of correct identifications was calculated for each participant and then averaged across all participants for a given task.

4.5.2. *Electroencephalographic data*

EEG and EOG signals were analog filtered with a bandpass from 0.1 to 100 Hz (−3 dB cutoffs) and digitized at 250 Hz. Trials containing EEG or EOG activity exceeding 100 μV were rejected automatically. The EEG or EOG was averaged for each stimulus in each condition over epochs of 700 ms, using a 100-ms prestimulus interval as baseline. Prior to statistical analysis, each stimulus average was filtered digitally with a low pass of 40 Hz using a Blackman window (61 dB). Peak amplitude and peak latency were measured at each scalp location for five ERP components – P1 (48–100 ms), N1 (100–200 ms), P2 (140–240 ms), N2 (200–352), and P3 (248–500 ms) – to each stimulus in each task and dimension. In addition, a late slow wave was measured as the average ERP voltage 500–700 ms after stimulus onset. All main effects and interactions reported as significant were reliable after Greenhouse–Geisser correction (Greenhouse and Geisser, 1959). ERP recordings were analyzed only for correct responses.

Table 1 – Stimuli and tasks across talker and vowel dimensions

Talker dimension Question asked: Is this talker 1 or 2?		Vowel dimension Question asked: Is this vowel [ɛ] or [æ]?	
Task	Stimuli	Task	Stimuli
Baseline 1	Talker 1 [ɛ] Talker 2 [ɛ]	Baseline 1	Vowel [ɛ] Talker A Vowel [æ] Talker A
Baseline 2	Talker 1 [æ] Talker 2 [æ]	Baseline 2	Vowel [ɛ] Talker B Vowel [æ] Talker B
Filtering	Talker 1 [ɛ], [æ] Talker 2 [ɛ], [æ]	Filtering	Vowel [ɛ] Talker A, B Vowel [æ] Talker A, B
Correlated 1	Talker 1 [ɛ] Talker 2 [æ]	Correlated 1	Vowel [ɛ] Talker A Vowel [æ] Talker B
Correlated 2	Talker 1 [æ] Talker 2 [ɛ]	Correlated 2	Vowel [æ] Talker A Vowel [ɛ] Talker B

Acknowledgments

We would like to thank David Kemmerer for his helpful comments on earlier drafts of this paper. This research was supported in part by NIH-NIDCD grant 5R03DC006811.

REFERENCES

Allen, J.S., Miller, J.L., 2004. Listener sensitivity to individual talker differences in voice-onset-time. J. Acoust. Soc. Am. 115, 3171–3183.

Baddeley, A.D., 1986. Working memory. Oxford Univ. Press, Oxford, England.

Baddeley, A.D., 1990. Human memory: Theory and practice. Erlbaum, Hove, England.

Bashore, T.R., Ridderinkhof, K.R., 2002. Older age, traumatic brain injury, and cognitive slowing. Psychol. Bull. 128 (1), 151–198.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.

Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport 14, 2105–2109.

Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. 8, 551–565.

Calhoun, V.D., Adali, T., Pearlson, G.D., Kiehl, K.A., 2006. Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. NeuroImage 30, 544–553.

Caramazza, A., Chialant, D., Capasso, R., Miceli, G., 2000. Separable processing of consonants and vowels. Nature 403 (27), 428–430.

Dien, J., Spencer, K.M., Donchin, E., 2004. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41, 665–678.

Donchin, E., Coles, M.G.H., 1988. Is the P300 component a manifestation of context updating. Behav. Brain Sci. 11, 355–372.

Duncan-Johnson, C.C., 1981. P300 latency: a new metric of information processing. Psychophysiology 18, 207–215.

Eisner, F., McQueen, J.M., 2005. The specificity of perceptual learning in speech processing. Percept. Psychophys. 67 (2), 224–238.

Garner, W.R., 1974. The Processing of Information and Structure. Lawrence Erlbaum Associates, Potomac, MD.

Garner, W.R., Felfoldy, G.L., 1970. Integrality of stimulus dimensions in various types of information processing. Cogn. Psychol. 1, 225–241.

Ghazanfar, A.A., Hauser, M.D., 1999. The neuroethology of primate vocal communication: substrates for the evolution of speech. Trends Cogn. Sci. 3, 377–384.

Ghazanfar, A.A., Hauser, M.D., 2001. The auditory behavior of primates: a neuroethological perspective. Curr. Opin. Neurobiol. 11, 712–720.

Ghazanfar, A.A., Smith-Rohrberg, D., Hauser, M.D., 2001. The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. Brain Behav. Evol. 58, 163–172.

Green, K.P., Tomiak, G.R., Kuhl, P.K., 1997. The encoding of rate and talker information during phonetic perception. Percept. Psychophys. 59 (5), 675–692.

Greenhouse, E., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–102.

Hauser, M.D., 1997. The Evolution of Communication. The MIT Press, Cambridge, MA.

Heisz, J.J., Watter, S., Shedden, J.M., 2006. Progressive N170 habituation to unattended repeated faces. Vision Res. 46, 47–56.

Kanwisher, N., McDermott, J., Chun, M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for the perception of faces. J. Neurosci. 17, 4302–4311.

Klostermann, F., Wahl, M., Marzinzik, F., Schneider, G.H., Kupsch, A., Curio, G., 2006. Mental chronometry of target detection: human thalamus leads cortex. Brain 129, 923–931.

Knosche, T., Lattner, S., Maess, B., Schauer, M., Friederici, A.D., 2002. Early parallel processing of auditory word and voice information. NeuroImage 17, 1493–1503.

Kreiman, J., 1997. Listening to voices: theory and practice in voice perception research. In: Johnson, K., Mullennix, J.W. (Eds.), Talker Variability in Speech Processing. AP Academic Press, New York, pp. 85–108.

Kreiman, J., Vanlancker-Sidtis, D., Gerratt, B.R., 2005. Perception of voice quality. In: Pisoni, D.B., Remez, R.E. (Eds.), The Handbook of Speech Perception. Blackwell, Malden, MA, pp. 338–362.

Kutas, M., Van Petten, C.K., 1994. In: Gernsbacher, M.A. (Ed.), Psycholinguistics electrified: event-related brain potential investigations. In Handbook of Psycholinguistics. Academic Press, New York, pp. 83–143.

Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–795.

Lavner, Y., Gath, I., Rosenhouse, J., 2000. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun. 30, 9–26.

Leuthold, H., Sommer, W., 1998. Postperceptual effects and P300 latency. Psychophysiology 35, 34–46.

Levy, D.A., Granot, R., Bentin, S., 2003. Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40, 291–305.

Lew, H., Shmiel, R., Jerger, J., Pomerantz, J.R., Jerger, S., 1997. Electrophysiological indices of Stroop and Garner interference reveal linguistic influences on auditory and visual processing. J. Am. Acad. Audiol. 8, 104–118.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.

Lockhead, G.R., 1966. Effects of dimensional redundancy on visual discrimination. J. Exp. Psychol. 72, 95–104.

Lockhead, G., 1972. Processing dimensional stimuli: a note. Psychol. Rev. 79, 410–419.

Magliero, A., Bashore, T.R., Coles, M.G.H., Donchin, E., 1984. On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology 21, 171–186.

Makela, A.M., Alku, P., Tiitinen, H., 2003. The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. Neurosci. Lett. 353, 111–114.

Martin, R.C., Romani, C., 1994. Verbal working memory and sentence comprehension: a multiple-components view. Neuropsychology 8, 506–523.

McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. Science 211, 77–80.

Melara, R.D., Marks, L.E., 1990. Processes underlying dimensional interactions: correspondencies between linguistic and nonlinguistic dimensions. Mem. Cogn. 18, 477–495.

Melara, R.D., Marks, L.E., Potts, B.C., 1993. Early-holistic processing or dimensional similarity? J. Exp. Psychol. Hum. Percept. Perform 19, 1114–1120.

Miller, J.L., Dexter, E.R., 1988. Effects of speaking rate and lexical status on phonetic perception. J. Exp. Psychol. Hum. Percept. Perform 14 (3), 369–378.

Mullennix, J.W., Pisoni, D.B., 1990. Stimulus variability and processing dependencies in speech perception. Percept. Psychophys. 47, 379–390.

Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav. Brain Sci. 13, 201–288.

Novak, G.P., Ritter, W., Vaughan, H.G., Wiznitzer, M.L., 1990. Differentiation of negative event-related potentials in an auditory discrimination task. Electroencephalogr. Clin. Neurophysiol. 75, 255–275.

Nygaard, L.C., Pisoni, D.B., 1998. Talker-specific learning in speech perception. Percept. Psychophys. 60, 355–376.

Nygaard, L.C., Sommers, M.S., Pisoni, D., 1995. Effects of stimulus variability on perception and representation of spoken words in memory. Percept. Psychophys. 57, 989–1001.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Oostenveld, R., Praamstraa, P., Stegemana, D.F., van Oosteromb, A., 2001. Overlap of attention and movement-related activity in lateralized event-related brain potentials. Clin. Neurophysiol. 112, 477–484.

Picton, T.W., 1992. The P300 wave of the human event-related potential. J. Clin. Neurophysiol. 9, 456–479.

Polich, J., Kok, A., 1995. Cognitive and biological determinants of P300: an integrative review. Biol. Psychol. 41, 103–146.

Polich, J., Herbst, K.L., 2000. P300 as a clinical assay: rationale, evaluation, and findings. Int. J. Psychophysiol. 38, 3–19.

Pomerantz, J.R., Pristach, E.A., Carson, C.E., 1989. Attention and object perception. In: Shepp, B.E., Ballesteros, S. (Eds.), Object perception: Structure and process. Erlbaum, Hillsdale, NJ, pp. 53–89.

Pritchard, W.S., Shappell, S.A., Brandt, M.E., 1991. Psychophysiology of N200/N400: A review and classification scheme. In: Jennings, J.R., Ackles, P.K., Coles, M.G.H. (Eds.), Advances in psychophysiology, vol. 4. Jessica Kingsley Publishers, London, pp. 43–106.

Puce, A., Allison, T., Gore, J.C., McCarthy, G., 1995. Face-sensitive regions in human extrastriate cortex studied by functional MRI. J. Neurobiol. 74, 1192–1199.

Raphael, L.J., 2005. Acoustic cues to the perception of segmental phonemes. In: Pisoni, D.B., Remez, R.E. (Eds.), The Handbook of Speech Perception. Blackwell Publishing, Malden, MA, pp. 182–206.

Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D., 1981. Speech perception without traditional speech cues. Science 212, 947–950.

Ritter, W., Simson, R., Vaughan Jr., H.G., Friedman, D., 1979. A brain event related to the making of a sensory discrimination. Science 203, 1358–1361.

Robinson, K., Patterson, R.D., 1995. The stimulus duration required to identify vowels, their octave, and their pitch chroma. J. Acoust. Soc. Am. 98 (4), 1858–1865.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R.J., Näätänen, R., 2003. Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. NeuroReport 13 (7), 1813–1816.

Shtyrov, Y., Pihko, E., Pulvermuller, F., 2005. Determinants of dominance: is language laterality explained by physical or linguistic features of speech? NeuroImage 27, 37–47.

Smith, L.B., Kemler, D.G., 1978. Levels of experienced dimensionality in children and adults. Cogn. Psychol. 10, 502–532.

Smith, L.B., Kilroy, M.C., 1979. A continuum of dimensional separability. Percept. Psychophys. 25 (4), 285–291.

Stevens, K.N., 2005. Features in speech perception and lexical access. In: Pisoni, D.B., Remez, R.E. (Eds.), The Handbook of Speech Perception. Blackwell Publishing, Malden, MA, pp. 125–155.

Swartz, B.L., 1992. Gender difference in voice onset time. Percept. Mot. Skills 75, 983–992.

Tervaniemi, M., Medvedev, S.V., Alho, K., Pakhomov, S.V., Roudas, M.S., vanZuijen, T.L., Näätänen, R., 2000. Lateralized automatic auditory processing of phonetic versus musical information: a PET study. Hum. Brain Mapp. 10, 74–79.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., Livingstone, M.S., 2006. A cortical region consisting entirely of face-selective cells. Science 311, 670–674.

Verleger, R., 1997. On the utility of P3 latency as an index of mental chronometry. Psychophysiology 31, 359–369.

von Kriegstein, K., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NeuroImage 22, 948–955.

von Kriegstein, K., Eger, E., Kleinschmidt, A., Giraud, A.L., 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. Cogn. Brain Res. 17, 48–55.

Wood, C.C., Day, R.S., 1975. Failure of selective attention to phonetic segments in consonant-vowel syllables. Percept. Psychophys. 17 (4), 346–350.

Xu, Y., Liu, J., Kanwisher, N., 2005. The M170 is selective for faces, not for expertise. Neuropsychologia 43, 588–597.