Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 35 (2007) 85–103
Journal of Phonetics
www.elsevier.com/locate/phonetics

The effect of talker familiarity on stream segregation

Rochelle S. Newman[a,*], Shannon Evers[b]

[a]*Department of Hearing & Speech Sciences, Program in Neuroscience & Cognitive Science, University of Maryland, College Park, MD 20742, USA*
[b]*Department of Psychology, E11 Seashore Hall, University of Iowa, Iowa City, IA 52242, USA*

Received 18 June 2004; received in revised form 30 October 2005; accepted 31 October 2005

Abstract

This study contrasts different forms of familiarity with a talker's voice, to better explore how these types of familiarity might influence a listener's ability to understand that voice in the context of noise. Listeners were asked to shadow a target voice while a second voice spoke fluently in the background. Listeners differed in their familiarity with the target voice: one group was familiar with the voice and were told explicitly whose voice they would be hearing (explicit knowledge); a second group was familiar with the target voice but were not warned whose voice it was (implicit knowledge), and members of a third group were entirely unfamiliar with the target voice. Explicit knowledge of talker identity appeared to have the larger effect on listener performance: Participants with explicit knowledge made significantly fewer shadowing errors than those with only implicit familiarity. Familiarity did influence the types of errors listeners made: those participants who were familiar with the target voice prior to the test session made fewer incorrect responses than did those who had not heard the talker previously, although their total number of errors (including misses, as well as incorrect responses) did not differ. In a second experiment, neither explicit knowledge of a distracter voice's identity nor implicit familiarity with the background speaker had any effect on shadowing a target voice. This suggests that familiarity with a voice only helps listeners when that voice is the one being attended. These studies suggest that there may be more than one manner in which talker-specific information influences perception.

1. Introduction

In a typical day, much of the speech we hear is likely to be from coworkers, friends, and family members: voices we know fairly well. Yet most research on speech perception has taken place in laboratory settings, with voices that were unknown to the participants prior to the start of the experiment. Listeners can understand these novel talkers quite well, and research on this ability is clearly relevant to understanding speech perception and phonetic identification generally. Yet understanding how familiarity with a talker influences perception is also an important aspect of speech perception, and one that has received much less attention.

Similarly, much of the speech we typically hear is likely to be in situations where there are multiple talkers speaking simultaneously. This poses a potential problem for perception, because unless listeners can separate

*Corresponding author. Tel.: +1 301 405 4226; fax: +1 301 314 2023.
E-mail address: rnewman@hesp.umd.edu (R.S. Newman).

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2005.10.004

the speech of the person to whom they are listening from the various background talkers, they are unlikely to be able to understand the spoken message.

The current research attempts to link these two domains of inquiry, by examining how familiarity with a talker's voice might influence a listener's ability to understand that voice in the context of speech from another talker. The following sections examine what is known about these two issues (talker familiarity and stream segregation) separately.

1.1. Effects of talker familiarity and talker identity

Recent years have seen an increase in the number of studies investigating influences of talker identity on speech perception. These studies have approached the issue from multiple angles. One approach has been to explore effects of talker variability: Researchers have compared listening conditions where only one person speaks with conditions in which multiple people speak in alternation (see for example Goldinger, Pisoni, & Logan, 1991; Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989; Nygaard, Sommers, & Pisoni, 1995; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1990, 1992). Since different talkers produce the same phonemes slightly differently, the latter condition forces the listener to frequently adjust their perception to a new talker (Kakehi, 1992), resulting in poorer perception overall, at least when the speaking rate is relatively fast. (With a slow speaking rate, listeners appear able to successfully encode talker-specific information, which they can later use to improve recognition memory; Goldinger et al., 1991.)

A second approach has been to compare performance for words spoken in familiar vs. unfamiliar voices. These studies have generally found processing advantages for words spoken in familiar voices. For example, recognition memory for a word is improved when it is presented in the same voice as it had been originally (Craik & Kirsner, 1974; Palmeri et al., 1993; Pisoni, 1990; Sheffert & Fowler, 1995; see also Church & Schacter, 1994). This is the case whether those words were actually heard in the particular voice, or simply imagined as having occurred in that voice (Geiselman & Glenny, 1977). These studies suggest that talker information is always part of the long-term representation of words (but see Geiselman & Bellezza, 1977).

Thus, voices that are familiar to the listener may lead to different processing or enhanced perception in a variety of tasks. But there are many degrees of familiarity one can have with a voice; familiarity in the studies described above generally consisted of having heard that voice on a number of previous words in the same experiment. This is likely to be quite different from the degree of familiarity that comes from interacting with an individual on a daily basis. Presumably, the familiarity one has with the voice of a friend or family member may be different from the short-term familiarity induced in a laboratory experiment. These different degrees of familiarity are generally not well distinguished in the literature, and most work has used relatively unfamiliar talkers.

There is some work that has investigated how familiar voices are recognized as familiar and identified, and what information might go into this process. Sheffert and colleagues (Sheffert, Pisoni, Fellowes, & Remez, 2002) found that listeners were able to learn to identify voices even when given only sinewave analogs to speech, which eliminate normal vocal quality while preserving phonetic cues and some (reduced) cues to vocal tract length. Listeners similarly learned voices from backward speech, which maintains vocal quality but distorts most phonetic cues. Abberton and Fourcin (1978) demonstrated that voices could be identified based solely on fundamental-frequency pattern characteristics alone. These findings suggest that listeners can make use of a wide variety of different acoustic cues to identify well-known voices. Moreover, work by Van Lancker and Kreiman suggests that listeners rely on different cues for recognizing different voices (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985). As a result, identification of some well-known voices is affected by speaking rate alterations where identification of others is not (Van Lancker, Kreiman, & Wickens, 1985), and some voices can be recognized backwards where others cannot (Van Lancker, Kreiman, & Emmorey, 1985). There appears to be no absolute set of cues that is used for all voices.

A few studies have begun investigating how familiarity with a voice can influence understanding (or processing) of speech in that voice. Schweinberger and colleagues (Schweinberger, Herholz, & Stief, 1997) found that listeners showed repetition priming for famous voices, but not for unfamiliar ones; when asked to decide whether a voice belonged to a famous individual, listeners performed better when they had been previously presented with another voice sample from that talker. They showed a smaller degree of priming

when they had been previously presented with that individual's face, and no priming when they had been previously presented with that individual's name. The fact that priming was greater for the famous talker's voice than for the talker's face or name suggests that famous voices activate a voice representation for the talker which is to some degree independent from other information about that talker.

Nygaard and colleagues (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994) found more accurate speech perception for a trained voice than for a novel voice. The authors trained a group of undergraduates to associate 10 voices with different names (for example, to learn that one voice was Peter, but another voice was Sam). Those listeners who succeeded at this later showed better speech identification in noise when the speech was from these known talkers. The authors suggested that perceptual learning of a voice influences later intelligibility, and thus that the stored information about a voice is used in (and helpful for) the "perceptual processing of the phonetic content of that speaker's novel utterances" (Nygaard et al., 1994, p. 44). Magnuson, Yamada, and Nusbaum (1994a, b, 1995) used participants' own spouses and children as experimental voices, and found that listeners were faster and more accurate at identifying speech in a well-known voice, especially when the speech was presented in noise. However, this advantage for familiar voices was apparently unrelated to effects of talker variability. Frequent alternation among a set of talkers was as disruptive to listeners when all of the voices were well-known as when the voices were all novel. This suggests that talker familiarity and talker variability have separate, additive effects on perception.

One area that has received less attention is the relationship between being familiar with a voice (having implicit knowledge of the talker's voice characteristics) and being able to identify it (knowing explicitly who is doing the talking). In most real-world situations, these two types of knowledge are confounded: when you hear a familiar voice, you often see the talker, and that visual stimulus provides explicit knowledge about the talker's identity. Yet there are situations in which these types of knowledge are uncoupled, such as with telephone conversations. Many individuals anecdotally report having the experience of answering the telephone and beginning a conversation without being positive to whom they are speaking: Often, the voice itself sounds familiar, but the specific identity of the talker remains unclear. This suggests that we do occasionally hear speech from a familiar talker without having explicit awareness of the talker's identity. In fact, experimental tests have suggested that when presented with voices of famous people, listeners often identify the voices as sounding familiar without being able to provide any identifying information about the individual (Hanley, Smith, & Hadfield, 1998).

Yet these two forms of talker familiarity have rarely been separated experimentally. In fact, it is often unclear which form of familiarity is being tested in a given study. For example, Nygaard and Pisoni (1998) trained listeners on 10 voices over 10 sessions, and then presented the listeners with an intelligibility test. They compared listeners' performance for these known voices with performance for unknown voices. But the authors do not report whether the participants were told they would be hearing the same voices learned earlier. We cannot be sure whether the listeners knew explicitly who was talking or were merely familiar with the voices (although we can presume that after 10 previous training sessions in the same laboratory setting, listeners probably suspected they would be hearing the same voices again, even if they were not explicitly told so.)

Yonan and Sommers (2000) attempted to distinguish between explicit and implicit knowledge through the use of different participant age groups. They found that elderly listeners were less accurate at identifying that a talker was familiar to them, but showed the same (or even a greater) benefit of voice familiarity in an identification task as did younger listeners. The authors take this to suggest that effects of voice familiarity are, at least in part, the result of implicit knowledge, rather than explicit knowledge. However, it is not entirely clear whether these elderly listeners had increased intelligibility even for those voices that they had failed to recognize (in which case only implicit knowledge was available), or if they only showed the voice familiarity effect on those voices for whom there was explicit knowledge as well. Although the comparable size of the familiarity effect across the age groups suggests that the effect is not limited to the voices explicitly recognized (as there are less of these for the elderly group), the difference in the number of voices recognized could be outweighed by a greater reliance on talker information by elderly listeners. Thus we still cannot be certain whether familiarity can influence perception in the absence of voice recognition.

Schweinberger, Herholz, and Sommer (1997) suggested that hearing a well-known voice would activate a stored perceptual representation of the voice; this activation could then spread to semantic information about

the talker. However, in some circumstances, this spreading of activation might be unsuccessful, such that the voice would elicit a sense of familiarity without any specific information about the talker becoming available. Using a variant of the gating task, they presented listeners with successively longer portions of speech from famous and unfamiliar voices, and asked listeners to explicitly identify the talker. They found that the greatest increases in the ability to recognize a voice as being familiar occurred within the first second of hearing that voice. When listeners failed to identify the voice, additional samples of the voice generally did not help; the authors took this as suggesting that different types of information were most relevant to accessing these different types of talker information (familiarity vs. explicit identification).

Yet this study did not investigate the effects these different types of knowledge might have on perception of speech from those talkers. Most models of talker effects assume that a familiar voice will activate voice-specific information, making the task of identifying the linguistic message easier. Often this is presumed to occur fairly automatically, suggesting that it is the familiarity with a talker's speech, not explicit (i.e., conscious) identification of the talker, that is the crucial aspect. Indeed, Craik and Kirsner (1974) have argued that effects of talker variability are based entirely on physical aspects of the stimuli, and that identity of a talker does not simply serve as a retrieval cue. Likewise Goldinger (1998) has argued that talker effects occur through automatic activation of episodic traces in the lexicon. These models thus focus on familiarity with a voice itself, and have little role for effects of talker identification. Yet explicit knowledge of a talker's identity might also have an influence on perception, especially if that allows the listener to activate an appropriate set of talker-specific phonological representations. Given these theoretical issues, there would appear to be a need for a more nuanced consideration of the notion of familiarity.

1.2. Effects of stream segregation

Speech often occurs in noisy environments, and in these situations the energy reaching the ear consists of information from multiple sources. Listeners need to be able to separate this amalgamation into original constituents in order to interpret the speech signal. This process (known as "streaming") appears to be based on a wide variety of acoustic cues, including perceived spatial location, talker sex, and voice pitch/fundamental frequency (e.g., Broadbent, 1952, 1954; Brokx & Nooteboom, 1982; Cherry, 1953; Culling & Darwin, 1993; Hirsh, 1950; Pollack & Pickett, 1958; Poulton, 1953; Shackleton & Meddis, 1992; Spieth, Curtis, & Webster, 1954; Summerfield & Assman, 1991; Treisman, 1960). Temporal synchrony (Dannenbring & Bregman, 1978) and amplitude modulation (Grimault, Bacon, & Micheyl, 2002; Grose, Hall, & Mendoza, 1995) can also serve as cues. For nonspeech sounds, both basic cues such as frequency (Bregman & Campbell, 1971), and more complex acoustic cues, such as static and dynamic cues to timbre (Cusack & Roberts, 2000; Iverson, 1995), can be used for segregation, although only the more basic auditory cues have been shown to be directly relevant to the ability to segregate speech (Mackersie, 2003; Mackersie, Prida, & Stiles, 2001). There is also work suggesting that Gestalt properties, such as common fate (Bregman, Abramson, Doehring, & Darwin, 1985) may play a role in separating sound streams (see also Bregman, 1990), although here, too, there is debate over the extent to which such properties are relevant to speech, or only to simple auditory stimuli (see Remez, Rubin, Burns, Pardo, & Lang, 1994).

Many of the properties used in stream segregation have to do with basic physical (acoustic) dissimilarity between the sound streams (see also Hartmann & Johnson, 1991). Listeners generally find it easier to separate streams, and to maintain attention on a particular stream, when the streams have less acoustic similarity to one another. This poses particular problems for listening to speech in multi-talker environments. Speech typically contains energy at a wide range of frequencies (ranging from less than 200 Hz to over 8000 Hz, depending on the talker and the phonetic segment). Moreover, the frequency ranges used by different talkers overlap for most of this range. This makes it very difficult to separate different speech streams, as there is inherent acoustic similarity between the different source signals (see, for example, Brungart & Simpson, 2002). It also makes it more difficult to focus attention on a particular stream, since the streams are less easily discriminable.

Talker familiarity could potentially play a role in these processes of stream segregation and selective attention. Familiar talkers might capture or hold attention relative to novel talkers. This could potentially be related to enhanced speech processing in noisy environments. Similarly, explicit knowledge of a talker's

identity may provide listeners with a useful cue for grouping components together or for focusing attention on the correct voice.

However, the role of talker information in stream segregation and selective attention tasks has not received much attention. While there have been studies investigating the role of talker familiarity on speech perception in noise (see for example Nygaard & Pisoni, 1998; Nygaard et al., 1994), none of these have used other talkers as the background noise. Thus little is known as to whether information about a talker (of either the explicit or implicit form) can influence the ability to separate different streams of speech and attend selectively. This is especially surprising given debates over whether higher forms of knowledge can influence stream segregation, or whether segregation is primarily influenced by peripheral (i.e., acoustic) processes (Cusack, Deeks, Aikman, & Carlyon, 2004; Hartmann & Johnson, 1991).

A recent study with infants (Barker & Newman, 2004) suggests that they benefit from hearing a known voice in a multi-talker environment. Infants aged 7.5 months heard either their own mother or another infant's mother as the target voice, while a second female talker spoke fluently in the background. The infants were tested for later recognition of words spoken by the target voice. Those infants who heard their own mother performed significantly above chance, whereas those who heard a novel target voice did not. Although talker familiarity cannot be separated from motivational issues in this study, it is at least suggestive that familiarity with a talker's voice could improve stream segregation/selective attention performance in adults, as well.

1.3. The present study

Thus, to summarize, there are several reasons to suspect that familiarity with a talker's voice might influence the ability to perceive that speech in a noisy environment. Moreover, explicit recognition of a talker might have an influence on perception, over and above any effects of voice familiarity. The present study is designed to explore these possibilities, by comparing listening performance in multi-talker conditions in which a target voice is either familiar and explicitly identified (explicit-knowledge group), familiar but not explicitly identified (implicit-knowledge group), or previously unfamiliar (novel-voice group). Our aim is to determine which type of knowledge has the greater influence on perception, at least in one particular type of task, that of selective attention.

The present pair of studies serves as an exploratory investigation of these issues. In Experiment 1, three groups of listeners were instructed to shadow a target voice while a distracter voice spoke fluently in the background. The target voice was that of an Introductory Psychology professor. The participants were students taking Introductory Psychology in one of two semesters during that academic year. Some students were taking the course during the semester that this professor was teaching, in which case his voice would presumably be familiar to them. Other students were taking the course the alternate semester, in which a different professor was teaching the course. For these students, the target voice would be unknown.

The participants hearing their own professor were themselves split into two groups; half of the participants were warned ahead of time whose voice they would be listening to, while the other participants were not. These two groups differ in whether they were given *explicit* information as to the talker's identity or only had implicit knowledge of the talker. (It is worth noting that experimentally distinguishing between explicit and implicit knowledge is complicated, in part because those individuals who were not given explicit knowledge might still develop such knowledge during the course of the experiment; this point is discussed in more detail later.)

Unlike much of the previous work on voice familiarity, listeners in the present study were not trained on the target voice prior to the test session. Any familiarity with the target voice was entirely the result of incidental learning occurring in a real-world context. Furthermore, the listeners were not presented with this voice in any earlier laboratory session, and the professor was not associated with this particular research lab. Thus the students had no reason to expect to hear this particular voice (or any familiar voice) in the present study. This allows for a more natural assessment of talker familiarity effects than that found in most laboratory studies (although some recent work has suggested that the two lead to very similar results; see Sheffert et al., 2002).

To summarize, three groups of participants took part in the present experiment: listeners who were unfamiliar with the target voice, listeners who were familiar with the target voice but were not given explicit cues to his identity, and listeners who were both familiar with the target voice and aware of whose voice they

were hearing. We examine effects of *explicit knowledge* of talker by comparing the performance of those individuals who were told the identity of the talker with those who were familiar with the voice but were uninformed (explicit vs. implicit knowledge groups). We also examine the role of *talker familiarity*, by comparing those individuals who were taking a course from this professor with those who were not, where neither group were told to whom they would be listening (implicit knowledge vs. novel voice groups). Given prior work suggesting that listeners with explicit knowledge of a talker's identity perform better at recognizing speech by that talker in the context of white noise, we expect to likewise find an advantage in multi-talker babble. Thus, we expect that the explicit group should perform better than the novel-voice group. If explicit knowledge is the primary factor in talker familiarity effects, we would expect the implicit-knowledge group to perform similarly to the novel-voice group; if familiarity with the voice itself is the primary factor, we would expect the implicit-knowledge group to perform similarly to the explicit-knowledge group. If both types of familiarity play a role, we would expect the implicit-knowledge group to show a level of performance between those of the other two groups.

2. Experiment 1

2.1. Method

2.1.1. Participants

A total of 67 students participated in this study; they were assigned to one of three groups. Of these, 24 were taking Introductory Psychology from a faculty member other than the target speaker; they made up the novel voice group. As this course is a prerequisite for all other psychology courses in the department, these students would not have taken other classes from the target professor. The remaining 44 participants had the target speaker as their professor; they were tested late in the semester, and their data were excluded if they did not attend class regularly. Thus, we presume that the students would have had ample time to learn this talker's speaking style, and would be relatively familiar with his voice. Half were instructed whose voice they would hear (the explicit-knowledge group), and the others were not (the implicit-knowledge group). Although it is possible that some of the listeners in the implicit-knowledge group spontaneously recognized the voice they were listening to, most reported being unaware of the talker's identity until the experimenter raised the question. Moreover, any tendency on the part of these students to spontaneously identify the target voice would serve to make the two groups more alike. Thus, while we cannot make strong claims if these two groups show similar performance, any difference between the two groups would clearly be a result of explicit knowledge about the talker.

Data from 10 participants were lost as a result of experimenter error or equipment failure (generally, the tape ending prior to completion of recording; $n = 5$) or because they reported they either "rarely" or only "sometimes" attended class ($n = 5$), suggesting they might not have familiarity with their professor's voice. This left a total of 57 participants, of whom 19 were in the explicit knowledge group, 16 in the implicit knowledge group, and 22 in the novel voice group.

2.1.2. Stimuli

Participants were asked to shadow two speech passages, a fluent story and a list of isolated words. The word list presumably has the characteristic voice pitch and vocal tract properties of the target voice, but may not contain the prosodic characteristics typical of the talker's fluent speech. That is, while the word list and the fluent story are likely to share many acoustic properties (e.g. physiologically-related properties such as pitch range and formant ranges; aspects of voice quality; some dialectal or idiolectal features), they are likely to differ in their patterns of intonation and rhythm; lists of isolated words have their own rhythmic properties, different from those in connected speech, and do not contain the prosodic properties that typically occur across words in connected speech. If knowledge of a talker's speaking style consists largely of those acoustic properties that are shared, we would expect familiarity effects to be the same for the word list and the story passage. If knowledge of a talker's voice draws crucially on prosodic properties such as rhythm, we would expect effects of familiarity only for the story passage, and not for the word list. Indeed, studies in which listeners are explicitly trained on a target voice have shown that familiarity effects do not transfer from fluent

speech to word lists (that is, when listeners have learned a voice from experience with fluent speech, they do not demonstrate recognition of that voice when tested with isolated words; Nygaard & Pisoni, 1998).

The target speaker was asked to read both the word list and the story passage in a style typical of how he spoke when teaching, although the accuracy with which he did so was not directly assessed. All recordings were made in a sound-attenuated chamber, using a Shure SM51 microphone, at a 44.1 kHz sampling rate. They were amplified, low-pass filtered, digitized via a 16-bit, analog-to-digital converter, and stored on computer disk. The word list consisted of monosyllabic words in English that had a familiarity rating of at least 6.0 on a 7-point scale (Nusbaum, Pisoni, & Davis, 1984) where 1 = completely unknown word, 4 = the word is recognized, but its meaning is unknown, and 7 = very well-known word. The story was taken from Louise Erdrich's "The Red Convertible." A second male talker recorded two stories to serve as distracters (or background passages) for the target story and word list. This second talker was a member of the laboratory, but did not test the participants in any of these sessions. None of the listeners reported being familiar with this background talker; nor did he recognize any of their names. Target and distracter passages were blended together binaurally, such that there were no spatial location differences, and such that the speech of the target voice averaged 5 dB more intense than the distracter passage; this level was chosen as a reasonable level of difficulty on the basis of pilot results.

The target voice spoke in isolation for the first 1 min of the passage to allow participants a chance to become skilled at shadowing his voice before a distracter was presented. This initial period allowed us to instruct listeners on which voice to follow without explicitly naming it. More critically, we were concerned that participants in the novel-group might become confused as to which voice they were supposed to follow without such familiarization. However, this necessarily gave the listeners time to become at least somewhat familiar with the voice they were supposed to be attending. This familiarization could reduce any differences across conditions, making it less likely that we would find effects. Thus, in some sense, all the listeners in this study were minimally familiar with the target speaker's voice, since they had heard him for at least 1 min prior to actual testing. However, the listeners still differed in the amount of such familiarization, since some had heard this talker for many hours in the classroom setting, in addition to this 1-min period. Still, the comparison on the basis of talker familiarity may be best thought of as a comparison between hearing a highly familiar and a slightly familiar voice, rather than a familiar vs. truly novel voice.

Participants were informed that the target voice would initially speak in isolation, but that a second voice would begin speaking part-way through; they were instructed to continue to shadow the target voice and ignore the distracter. No participant reported being confused as to the voice they were supposed to be attending, and no participant switched to attending the wrong voice during the experiment (although some participants did repeat isolated words from the wrong voice, as has been found in other such studies; Treisman, 1960).

In addition to these test passages, two 45-s practice passages (one word list, and one story) were created to ensure that the listeners were familiar with the task. Unfamiliar female voices were used for both practice passages.

2.1.3. Procedure

Participants were tested in a sound-attenuating chamber. Passages were presented over circumaural headphones at a comfortable listening level; circumaural headphones surround the ear, and thus partially block outside noises. This helps ensure that the participant's own voice did not mask that of the target speaker.

Half of the participants in each group heard the word list first, and the other half heard the story first. Before each portion, they were given a practice passage of the same type to shadow. Participants were instructed that they would be shadowing a talker's voice while another person spoke fluently in the background. That is, they were supposed to repeat back everything that the talker said, as that talker continued to speak. They were told that the target speaker would talk continuously, and thus that they needed to attempt to keep up with the talker. They were not given any specific instructions as to how closely in time they should shadow the message. Participants in the explicit knowledge-group were told prior to the target passage that the voice they would be shadowing was that of their Introductory Psychology professor. Participants in the other groups were not given this information.

After the final passage, participants were given a questionnaire, asking a number of questions. The first was whether the voices sounded familiar or whether they could identify any of the voices. Some of the participants in the implicit knowledge group commented that while they had not consciously recognized the target voice during the experiment, they were able to guess the identity of the talker when the questionnaire asked them to do so (indeed, 12 participants in this group identified the target voice; no subject identified the background voice). These comments imply that the questionnaire did not provide an accurate picture of participants' knowledge of the talker at the time of the recording. In fact, many of the participants did successfully guess the voice when asked to do so, although we have no way of telling whether they actually identified it during the experiment. Furthermore, even if listeners did identify the talker part-way through the experiment, we would have no way of knowing whether that identification occurred early or late in the passage. We therefore decided not to exclude data on the basis of successful voice recognition; if anything, this serves to decrease the likelihood that the implicit knowledge group would be different from the explicit knowledge group.

The questionnaire also asked how often they attended class. As class attendance was not taken, we have no way of verifying the accuracy of the participants' responses to this question. However, any overestimation of class attendance would serve to reduce listeners' familiarity with the target voice, and thus reduce the likelihood of any significant differences between groups, provided that participants in the three groups overestimate to the same extent. We excluded data from participants who admitted to attending their class "rarely" or only "sometimes", as they may not have had much exposure to the target voice. (Participants in the novel-voice group were not excluded for this reason, as this would not influence their knowledge of the talker.) Of the remaining participants in the two knowledge groups, 7 reported "always" attending class, 17 "almost always", 8 "usually", and 2 "half the time", with responses evenly spread across knowledge groups (explicit knowledge: 5 always, 9 almost always, 5 usually, 1 half). We also asked the students to indicate how often they attended class but found themselves not paying attention; 15 reported "rarely", 12 "occasionally", and 7 "sometimes, again with responses evenly spread across knowledge groups (explicit familiarity: 8 rarely, 8 occasionally, 3 sometimes).

Finally, the questionnaire asked participants to rate how distinctive their professor's voice is, and how likely they felt they would be to recognize it if presented. The only participant to rate themselves as being unlikely to recognize their professor's voice (<3 on the 5-point scale) only rarely attended class and was thus excluded; of those who attended class regularly, only one participant rated the distinctiveness of their professor's voice as below a 4 (average was 4.6). Average rating for being able to recognize the voice if heard was 4.2.

2.1.4. Scoring

The participants' speech was recorded using a Shure SM81 microphone routed to a 3-head Marantz tape deck. An experimenter listened to the tape of each participant's shadowing and compared this with a typewritten transcript of the actual passage. Errors consisted of any word that was incorrect, omitted, or inserted; the three types of errors were tabulated separately. We did not tabulate degrees of incorrectness; thus, a word that was off-target by multiple phonemes was considered "wrong" to the same extent as a word that had a one-phoneme substitution or addition. Across the two experiments reported here, data from 12 participants (10%) were re-coded by a second experimenter. Correlations between coders were .94 for total errors on the word list, and .99 for the story.

2.2. Results and discussion

Errors were analyzed using a two-way ANOVA in which each of the IVs (error type and knowledge type) had three levels. There was a significant effect of error type, $F(2, 108) = 126.27$ for stories, $F(2, 108) = 173.43$ for word lists, both $p < .0001$. Not surprisingly, most errors consisted of missed words, rather than wrong words or insertions. For the stories, there were, on average, 341 missed words, 83 wrong words, and 20 insertions. For the word lists there were 133 misses, 77 wrong words, and 2 insertions. These differences were so pronounced for the stories that the total number of errors was mainly driven by missed words. For this reason, although our primary focus was on total errors, we also examined errors by types, in order to determine whether there were effects in the wrong words or in the insertions that were overshadowed by misses.

There was a marginal effect of knowledge in the stories only (for stories, $F(2, 54) = 2.51$, $p < .10$; for lists, $F(2, 54) = 1.58$, $p > .20$). However, because we were particularly concerned with comparisons between the explicit and implicit knowledge groups, and between participants who were familiar and unfamiliar with the target voice, we had planned to make these comparisons regardless of the results of the overall analysis. As there were directional predictions, we used 1-tailed tests. We found that the effect of explicit knowledge did have a significant effect on listener performance in the story condition (explicit knowledge vs. implicit knowledge, $t(33) = 2.25$, $p < .05$). Listeners who were told who they would be hearing averaged 371 errors, while those who were not told (but were still familiar with the voice—the implicit-knowledge group) averaged 519 errors, as shown in Fig. 1. This effect is striking, given that some of the listeners who were not told may very well have identified the voice on their own. This accuracy difference is primarily from the number of missed words; the two groups had similar numbers of wrong words (74 vs. 81) and insertions (20 vs. 18); the difference in missed words was the only significant difference (270 missed words for those with explicit knowledge; 426 for those without; $t(33) = 2.40$, $p < .05$).

We next examined the effect of familiarity, by comparing those participants who were familiar with the target's voice (but were not told his identity; i.e., the implicit-knowledge group) with those who were not familiar with his voice (the novel-voice group), but we found no significant effects ($t(36) = .91$, $p > .10$). Looking only at the missed words, where the largest differences were to be found, again showed no significant effect ($t(36) = 1.17$, $p > .10$). Apparently, then, simple familiarity with a voice is not sufficient to aid in the process of stream segregation. In contrast, explicit knowledge of the talker's identity does appear to have an influence. This suggests that some of the effects of talker familiarity in previous research may have been the result of listeners having explicit access to information about the talker or talkers.

One possibility is that listeners in the implicit knowledge group were actually distracted by not knowing the identity of a seemingly familiar voice. Perhaps the members of this group were spending mental energy attempting to identify the talker, using up cognitive resources and thus preventing these resources from being applied to the streaming task. Rather than explicit knowledge being a benefit, familiarity without explicit knowledge would then be a hindrance. Although we cannot entirely rule out this possibility, such an explanation would actually suggest that listeners in the implicit knowledge group would do more poorly than participants with no experience with the voice (who presumably would have no ground for being distracted in the manner suggested). Looking at Fig. 1, it does appear as if this might be the case, since listeners in the

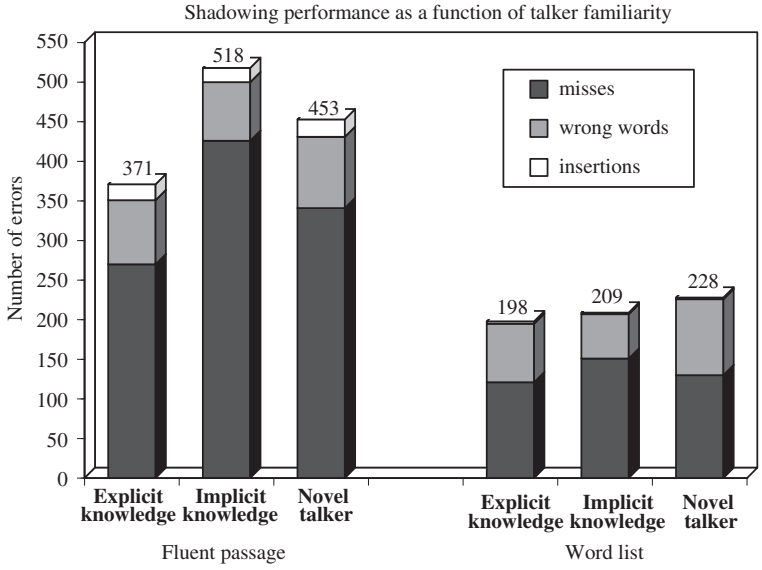

Fig. 1. Total number of errors, and errors of each type, for participants in Experiment 1 who were told the identity of the talker (explicit knowledge), who were familiar with the talker's voice but were not told (implicit knowledge) and for those who had no familiarity with the talker (novel voice). Results from the fluent speech passage are on the left; those from the word list are on the right.

novel-voice group made fewer misses than did individuals in the implicit knowledge group. However, this difference did not approach significance (as shown above), suggesting that the listeners were not simply distracted by attempts to identify the talker. That said, when we compare the performance of the explicit knowledge group to the novel voice group, we find only a marginal difference (for all words, $t(39) = 1.53$, $p < .07$; for missed words, $t(39) = 1.34$, $p < .10$). It is unclear whether this lack of significance is simply the result of the variability in these participants' performance, or is another indication that the novel-voice group performed less poorly than the implicit-knowledge group, but such a finding does suggest that a portion of the difference between the explicit and implicit knowledge groups might be distraction-based. That is, it may be that the present results demonstrate both an advantage for explicit knowledge and a disadvantage to devoting cognitive resources to talker identification, rather than simply an explicit-knowledge advantage.

Still, the data suggest that individuals who had explicit knowledge about the talker performed more accurately than those who had only implicit knowledge (and marginally more accurately than those in the novel-voice group), and that the implicit-knowledge group did not differ even marginally from the novel-voice group. This pattern suggests that the effect is primarily one of explicit knowledge improving performance, although the findings are clearly somewhat weak.

We next examined the data from the word lists, but here found no significant effects (for explicit knowledge vs. implicit knowledge, $t(35) = .70$; for implicit knowledge vs. novel voice, $t(36) = 1.00$; both $p > .10$; for explicit knowledge vs. novel voice, $t(39) = 1.63$, $p < .10$), suggesting that the advantage of explicit knowledge in the story condition was likely to be the result of knowledge of the typical intonational and rhythmic patterns in the talker's fluent speech, as compared to segmental, vowel quality, or pitch patterns. Although the word lists no doubt shared a great many acoustic properties with the stories (such as physiological properties of the speaker's voice), these characteristics were apparently not sufficient to provide a benefit to listeners in the present shadowing task. Or, perhaps the difference in intonational patterns between the connected speech in the classroom and the isolated words presented here prevented the listener's from benefiting from their knowledge of the talker's identity. This could suggest that a large component of learning a talker's voice is learning the rhythmic and intonational properties typical of that voice. Alternatively, it may be that learning a voice is to some extent context-specific: since these listeners had (presumably) learned this voice while he was speaking in complete sentences, that knowledge simply did not generalize to other types of speech passages. Although the present study does not discriminate between these two explanations, recent work by Nygaard and colleagues (Nygaard, Burt, & Queen, 2000; Nygaard & Pisoni, 1998; Sheffert et al., 2002; see also Yonan & Sommers, 2000) suggests the latter is the more appropriate explanation. Thus, the present null effect for word lists is not necessarily the result of listeners only learning prosodic information about our talker, but may instead be the result of listeners failing to generalize knowledge about talkers to different contexts.

Having implicit knowledge clearly did not influence the total number of errors our listeners made. However, it did appear to have influenced the *types* of their errors. In addition to the main effects reported above, the overall ANOVAs showed significant interactions between type of error and level of knowledge for both the word lists ($F(4, 108) = 3.44$, $p < .05$) and the stories ($F(4, 108) = 2.87$, $p < .05$). (These results also hold when based on proportions, rather than absolute numbers.) Collapsing across the two types of passages, we find a significant effect of knowledge for both the number ($F(2, 54) = 5.87$, $p < .05$) and proportion ($F(2, 54) = 5.25$, $p < .05$) of wrong words; listeners in the explicit-knowledge group said 155 wrong words, those in the implicit-knowledge group said 130 wrong words, but those who were not familiar with the voice made 185 such errors. Follow-up t-tests showed that the novel-voice group made a greater number of wrong errors than both the explicit-knowledge group ($t(39) = 1.95$, $p < .05$) and the implicit-knowledge group ($t(36) = 3.84$, $p < .0005$), but that the two familiar groups did not differ ($t(33) = 1.33$, $p > .10$). Thus, listeners seemed more likely to say a word incorrectly when they were not familiar with the voice that was speaking.

In summary, it appears that listeners miss fewer words when they know explicitly who is talking. In contrast, simple familiarity with a voice does not seem to have the same effect. Instead, those who were familiar with the voice appeared to make fewer wrong responses, even though they made no fewer errors overall.

One possibility is that missing words is related more directly with failure at following a voice, whereas saying a word incorrectly is more tied to issues of accurate perceptual identification. If so, this would imply that explicit knowledge helps listeners attend to a particular voice, whereas familiarity may help with

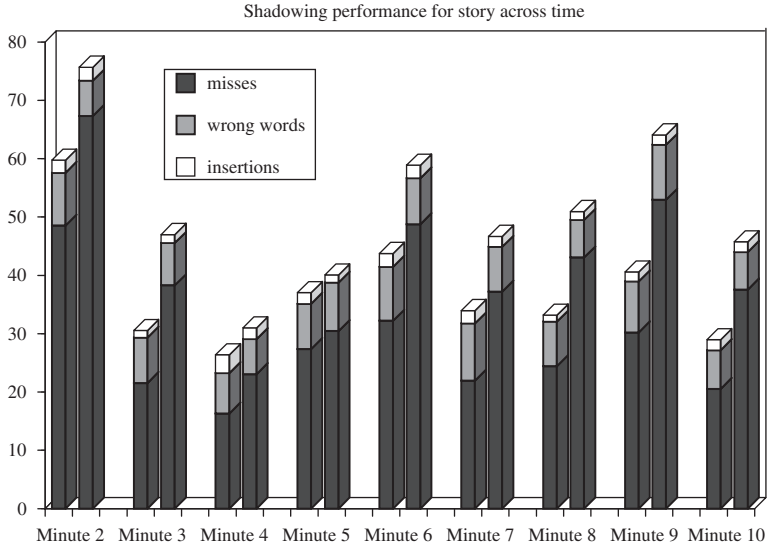

Fig. 2. Total number of errors, and errors of each type, for each minute of time. Results from participants who were told the identity of the talker (explicit knowledge) are in the left columns for each pair; results from participants who were familiar with the talker's voice but were not told (implicit knowledge) are in the right column for each pair.

interpreting the speech appropriately (rather than mishearing) once it has been attended. Yet the interpretation of different types of errors is a complicated issue. Familiarity with a voice could influence participants' response biases, rather than their actual sensitivity: perhaps listeners in the implicit-knowledge group were more conservative in their guessing than those in the novel group, being less likely to guess what a word was when they were unsure—this would result in a greater number of misses, but fewer wrong responses.[1]

One final question regards the time-course of this explicit-knowledge effect. Listeners in the implicit-knowledge group showed poorer performance overall than those individuals who were told whose voice to expect. However, it is possible that these listeners realized the identity of the speaker partway during the course of the experiment. If so, the difference between these two groups should be largest in the beginning of the experiment, but should decrease as the experiment continues. To evaluate this, we evaluated each minute of the story separately (see Fig. 2). Since our primary results were in misses, and there were relatively few wrong responses and insertions, we examined the missed words only.

The first minute of the passage occurred without the presence of background speech noise, and was therefore ignored. This left 9 additional 1-min time segments (with the final one of these being slightly less than a minute, 49.59 s). Across these nine segments, there was a significant overall effect of time ($F(8, 264) = 31.20$, $p < .0001$), a significant effect of knowledge group ($F(1, 33) = 5.72$, $p < .05$) and a significant time by knowledge group interaction ($F(8, 264) = 2.55$, $p < .05$). The explicit-knowledge and implicit-knowledge groups differed significantly in minutes two ($t(33) = 2.13$, $p < .05$), three ($t(33) = 2.49$, $p < .01$), six ($t(33) = 2.10$, $p < .05$), seven ($t(33) = 2.17$, $p < .05$), eight ($t(33) = 2.40$, $p < .05$), nine ($t(35) = 2.50$, $p < .01$), and ten ($t(33) = 2.88$, $p < .005$). There were no significant effects in minutes four ($t(33) = 1.40$, $p > .05$) or five ($t(33) = .45$, $p > .05$), although there was still a trend in the appropriate direction. Thus it does not appear that the effect of explicit knowledge differed between the beginning and ending of the experiment. The 2 min without significant effects (minutes 4 and 5) were two of the three minutes in which the novel voice group made the fewest number of misses, suggesting that these may have simply been easier portions of the story (minute 7 was tied with minute 5 for the 2nd fewest errors); the lack of a difference here may be the result of ceiling performance.

[1]We thank Rachel Smith for this suggestion.

Thus it appears that there are two ways in which knowledge about a talker can influence perception. Explicit information about who is speaking influences the ability to correctly follow and/or intepret a voice, as indicated by the reduced overall error rate (particularly the difference in missed words). Implicit familiarity with or knowledge about a voice influences either the ability to correctly perceive a voice once it has been heard, or the listener's response bias. The fact that the implicit familiarity effects were less consistent could be the result of the particular task used (one which particularly emphasized selective attention abilities, rather than acoustic discrimination). If so, we might potentially have the very opposite pattern (stronger effects of familiarity than of explicit knowledge) if the task were one of single-word identification in the midst of greater levels of noise.

The present effect of explicit knowledge could be caused by (at least) two different mechanisms. One possibility is that listeners are better able to separate the target voice from the background if they know whose voice it is; that is, talker knowledge may influence the ability to segregate different streams of speech *per se*. An alternative possibility is that the advantage of explicit knowledge is primarily attentional—rather than make the act of separating the voices easier, this knowledge makes it easier to attend to that voice after the act of separation is complete. If so, knowing the identity of the talker would reduce confusion as to which voice one should listen to, but would not affect earlier stages of processing such as stream segregation.

One way to distinguish between these two possibilities is to place the known voice in the background. If explicit knowledge about a talker's identity aids in *separating* the different streams of speech, rather than in the act of attending to one of them, we might expect to find the same advantage when the known voice was the one to be ignored. Which voice is later attended should in some sense be irrelevant to the act of separating the streams of speech from one another. In contrast, if the advantage of explicit knowledge was primarily one of making that stream of speech easier to attend, we would expect no such advantage of familiarity when the familiar voice is not the one being attended.

It is also quite possible that the acts of segregation and attention are not fully separable in this manner. Indeed, if stream segregation is itself an attentional act (or is in some way influenced by the act of selective attention; see Cusack et al., 2004), this division into two tasks is certainly overly simplistic. There have been a number of studies attempting to explore the extent to which an individual who is shadowing a particular message is aware of information in a concurrent stream (e.g., Broadbent, 1952; Cherry, 1953; Moray, 1959; Ostry, Moray, & Marks, 1976; Treisman, 1960; Wood & Cowan, 1995) or can perform additional tasks concurrently (see, for example, Allport, Antonis, & Reynolds, 1972). Other studies have examined selective attention through the use of implicit priming and electrophysiological responses, tasks that investigate whether individuals processed the alternative signal without being dependent on their memory for having done so (e.g., Bentin, Kutas, & Hillyard, 1995; Corteen & Wood, 1972; Dupoux, Kouider, & Mehler, 2003; Eich, 1984; MacKay, 1973; Rivenez, Darwin, & Guillaume, 2004; Wood, Stadler, & Cowan, 1997). These studies have often been motivated by competing models of selective attention (Broadbent, 1958; Cusack et al., 2004; Deutsch & Deutsch, 1963; MacKay, 1973). We do not wish to make claims for any specific model in the present paper, and certainly the division between selective attention and stream separation ignores many of the nuances inherent in attention research. However, by attempting to separate out the different aspects of stream segregation/selective attention in this manner, we can at least begin to gain an understanding of the ways in which knowledge influences attention. Experiment 2 is designed to investigate this issue, by having listeners attend to a novel voice, while the distracter voice was either a familiar talker or an unknown talker.

3. Experiment 2

This experiment was very similar to Experiment 1, except that the voice of the Introductory Psychology professor now served as the distracter voice.

3.1. Method

3.1.1. Participants

Sixty-five students participated in this experiment; they were assigned to one of three groups. Twenty-four of these participants were taking Introductory Psychology from a faculty member other than the distracter

speaker; they made up the novel voice group. The other participants had the distracter speaker as their professor; half were instructed whose voice they would be ignoring (the explicit knowledge group), and the others were not (the implicit knowledge group). It is important to note, however, that the target voice was novel to all three groups of participants. Data from 2 participants were dropped because the participants reported rarely attending class; data from 1 additional participant were excluded because the participant actually knew the target speaker. One participant did not complete the experiment, and data from another 3 were dropped for experimenter error or equipment failure. This left a total of 20 participants in each of the two familiar groups (explicit knowledge and implicit knowledge), and 18 participants in the novel-voice group.

3.1.2. Stimuli

A new male talker recorded both a story and a word list to serve as the target passages; the story came from *A Confederacy of Dunces* by John Kennedy Toole. The target story from Experiment 1 served as one of the distracter passages, and the Introductory Psychology professor recorded a second story to serve as the other distracter passage. The distracter passages were blended with the target passages such that the speech of the target voice averaged 10 dB more intense than the distracter passage; this value was chosen on the basis of pilot studies. As before, the target voice spoke in isolation for the first 1 min of the passage. This gave the listeners sufficient time to become familiar with the voice they were supposed to be attending. The same practice samples were used here as in Experiment 1.

3.1.3. Procedure and scoring

Both the procedure and scoring was identical to that in Experiment 1, and the testing took place during the same two semesters.

3.2. Results and discussion

Errors were again analyzed using a 2-way ANOVA with 3 levels of error type and 3 levels of knowledge. There was a significant effect of error type, $F(2, 110) = 375.44$ for stories, $F(2, 110) = 180.85$ for word lists, both $p < .0001$ As before, most errors consisted of missed words, rather than wrong words or insertions (for lists, there were 68 wrong words, 1 insertion, and 130 misses; for stories, there were 102 wrong words, 11 insertions, and 613 misses). There were no significant effects of knowledge or interactions between knowledge and error type (all $F < 1$). As in Experiment 1, we proceeded to our planned comparisons, but again found no significant differences. There was no difference between the explicit-knowledge and implicit-knowledge groups on either story performance ($t(38) = .05$, $p > .05$) or list performance ($t(38) = .78$, $p > .05$) using the total number of errors as the data; nor were there any differences between the implicit-knowledge and novel-voice groups ($t(36) = -.09$ for the story, and $t(36) = -.40$ for the list, both $p > .05$). In fact, performance was strikingly similar across groups, as shown in Fig. 3. Participants in the explicit-knowledge group made 729 errors in the story, and 203 in the list; those in the implicit knowledge group made 725 errors to the story and 191 to the list, and those in the novel voice group made 732 errors to the story and 198 to the list.

As noted above, the majority of errors were misses; these could potentially swamp any differences among groups of the other two error types. We therefore examined each type of error separately with 1-way ANOVAs with the factor of knowledge, as before, but found no significant difference (for misses and incorrect words, $F < 1$; for insertions, $F(2, 55) = 1.50$, $p > .10$). Proportions of errors also showed no significant effects, although those listeners who were unfamiliar with the talker made marginally fewer insertion errors than did those who were familiar (for the story, $t(36) = 1.73$, $p < .10$; for the list, $t(36) = 1.76$, $p < .10$, both two-tailed). This may be a hint that those familiar with the voice suffered more intrusions from that voice, but the marginal nature of the effect makes it impossible to place much weight on these differences.

It is worth noting that performance in this experiment was substantially poorer overall than that in Experiment 1, despite the fact that the target voice had a 10 dB advantage over the distractor (compared to only a 5 dB difference in Experiment 1). This could be a result of the particular combination of voices, or of this target voice being particularly difficult to follow. However, these differences between studies were most obvious in the story passages; there was actually very little difference in errors across the word lists in the two experiments. This suggests that the difference may actually be found in the stories themselves. In order to use

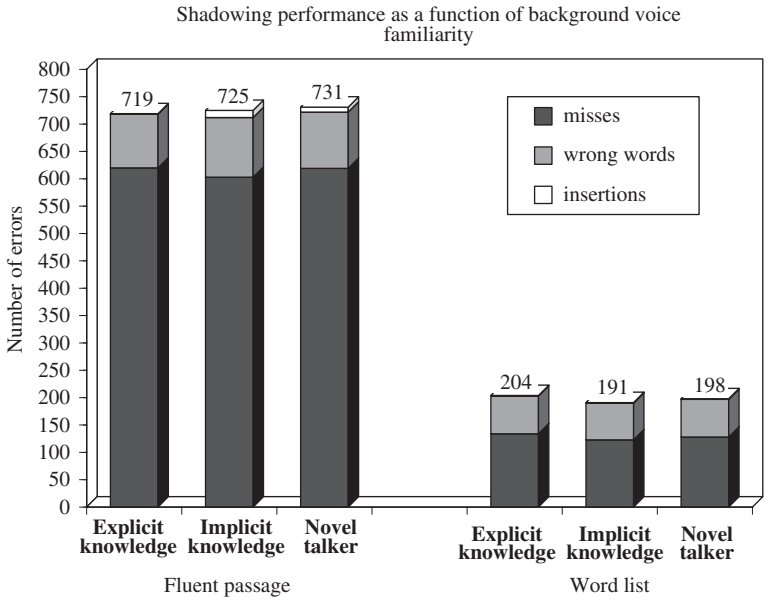

Fig. 3. Total number of errors, and errors of each type, for participants in Experiment 2 who were told the identity of the background talker (explicit knowledge), who were familiar with the background talker's voice but were not told (implicit knowledge) and for those who had no familiarity with the background talker (novel voice). Results from the fluent speech passage are on the left; those from the word list are on the right.

the same recording as target in Experiment 1, and distracter in Experiment 2, that story could not be the one shadowed in this experiment. In retrospect, the alternate story was in many ways less predictable from moment to moment than was that from Experiment 1, and this clearly made the task more difficult for the listeners. However, since the comparison in this experiment is between three groups hearing the same story, this difference should not influence the pattern of results, as long as performance was not at floor. Although individuals did average a large number of errors, the total number of words in the story was 1920; thus, listeners were still succeeding in correctly shadowing the majority of words, and performance does not appear to be at floor.

These results demonstrate that knowledge of a talker's identity can only influence perception if that talker is the one being attended, not if that talker is the one being ignored. In one sense, these results are not too surprising; Cherry (1953) found that listeners attending one voice were oblivious to many changes in the distracter voice, including changes from one language to another. If listeners generally fail to notice the language spoken by a distracter voice, the participants in the present study who were not informed of the talker's identity may likewise never even have noticed who was speaking. Indeed, many of them did not identify the voice as even being familiar when asked this on the questionnaire. However, one might have expected that those listeners who were told whom to ignore might have been able to use that information as a filter in order to help them pay attention to the other voice. This does not appear to be the case.

Similar results have been found in musical stream segregation. Dowling (1973) presented listeners with interleaved melodies, and found that listeners were better able to detect sequences with which they were already familiar, comparable to our findings in Experiment 1. However, familiarity with the background melody did not appear to help the listeners isolate the target melody. Bregman (1990, p. 407) interpreted these results in the following manner: "As we increase the difference in familiarity between two sound patterns that are mixed together by making ourselves more and more familiar with just one of them, although we may become increasingly more skillful at pulling the familiar one out of the mixture, this does not, in itself make it easier to hear the unfamiliar one as a coherent sequence." This seems to be an equally apt description of the present results with speech.

4. Discussion

There are two forms of talker familiarity that have often been confounded in the literature: familiarity with a talker's voice, and explicit knowledge of a talker's identity. The present experiment attempted to separate these different effects experimentally in the context of stream segregation. The results can be best thought of as a first attempt to begin addressing a very complex area of investigation. Moreover, the effects we found were relatively small. Despite these limitations, the present work clearly suggests that there is a need for future studies to consider the different forms of talker knowledge available to listeners.

We found that explicit knowledge appeared to have more substantial effects on perception than did voice familiarity. Listeners who were told the identity of the voice they were to attend made fewer errors than listeners who had equivalent familiarity but were not told the talker's identity. They also made fewer errors than listeners who had no knowledge of the talker at all—but this difference was only marginal statistically. Thus, while there does appear to be a pattern suggesting that explicit knowledge affects performance, the effect is not terribly strong.

In contrast, familiarity with a voice (implicit knowledge) did not appear to help listeners pay attention to that voice in a noisy environment. Listeners who were familiar with the talker made no fewer streaming errors than those who were not. Familiarity with a talker did have one effect on performance, however: those listeners who were familiar with the voice they were shadowing (but unaware of the talker's identity) appeared to do a more accurate job of interpreting that speech. They said significantly fewer incorrect words, even though their total number of errors did not differ. Thus, familiarity with a talker may influence the ability to accurately comprehend speech from a talker; alternatively, it may be that it changes the bias to respond, such that these listeners were less likely to guess when unsure than were listeners who knew explicitly to whom they listened. This type of change in response bias would result in a reduced number of wrong words, but also an increase in missed words, since the listener would simply opt not to respond unless certain of what was heard; this matches the pattern we found. Thus it remains unclear whether the implicit familiarity with a voice affects perception or response bias; in either case, it did appear to have an effect on the listeners' performance.

No advantages (in the form of reduced processing costs, leading to fewer errors) were found for a known voice when that voice was the one to be ignored. This could suggest that both the implicit effects of familiarity with a voice and that of explicit knowledge of a talker's identity have their influence at a stage of processing beyond that of "pure" stream segregation. However, it could also be that these forms of knowledge are processed early, but simply are unavailable to be exploited when the target voice is not the one being attended. Regardless of the processing explanation, it appears that only the act of attention to a given stream, not the act of segregating two streams, is influenced by talker-specific knowledge. This corresponds to previous findings for musical streaming (Dowling, 1973), and has some similarities to results by Hartmann and Johnson (1991) on listeners' ability to identify interleaved melodies. They found that peripheral channeling (based on simple acoustic information such as location in space and overall frequency range) could account for virtually all of their listeners' results. There was virtually no influence of more complicated grouping principles, such as temporal envelopes or rhythmic differences. Although Hartmann and Johnson did not examine truly "higher" forms of knowledge (such as being told the identity of the melody to listen for), the suggestion is that segregation itself is a very low-level process and would be unaffected by higher knowledge. The similarity between these previous findings and the present results suggest parallels between speech stream separation and musical stream segregation. While this may seem unsurprising, it has been an issue of some debate (Barker & Cooke, 1999). Remez and colleagues (Remez et al., 1994) have argued that the mechanisms responsible for grouping of speech signals cannot be the same as those used for more simple auditory stimuli (see Bregman, 1990), because phonemes that are extremely different acoustically (such as fricatives and vowels) nonetheless cohere as a stream, even though most auditory grouping principles would predict otherwise. Although the current research makes no attempt to explore the specific properties that cause a stream of speech to cohere, the finding that higher forms of knowledge do not have a strong influence supports the idea that segregation may occur in a fairly automatic (even modular) manner. Whatever the acoustic properties that influence speech grouping, they appear to be properties that can be specified through fairly simple rules (Barker & Cooke, 1999), or at least are "informationally encapsulated" (Fodor, 1983).

Bregman (1990) has also pointed to evidence separating effects of familiarity from effects of peripheral channeling. For example, increasing the speed of a musical melody results in stronger segregation on the basis of frequency. In contrast, the ability to select on the basis of familiarity becomes poorer with faster presentation rates (p. 669). Thus, while familiarity can certainly influence performance on streaming tasks, the mechanisms of such effects appear to be quite distinct from the mechanisms involved in primitive grouping.

Clearly, the results from this study need to be viewed as preliminary; the shadowing task is only one approach to investigating stream segregation abilities, and other tasks might be sensitive to different aspects of speech recognition processes. Moreover, individuals likely differ substantially in their familiarity with any given talker's voice, and studying voice familiarity requires that comparisons be performed in a between-subjects comparison, adding to this variability (that is, the same listener cannot be both familiar and unfamiliar with the same voice at the same time). Although it would be possible to compare the same listener's performance across different voices (one familiar and one unfamiliar), this approach raises its own set of difficulties, since the acoustic properties of the voices are left uncontrolled. Similarly, testing a listener with a voice and then providing explicit training is another method, but leads to a confound between task experience and voice familiarity. These facts place limitations on the sensitivity of any particular task investigating the effect of familiarity. There is thus a need to use multiple approaches to examine this factor, and the present study serves only as a first step in that direction.

Indeed, one possible explanation for the lack of an effect of implicit knowledge on selective attention may be related to the particular procedures employed in this experiment. This was a long experiment, involving 10 min of continuous speech, and the first 1 min of the target voice's speech was presented in isolation (without background noise). One minute could be long enough to achieve some aspects of familiarization. Thus, the novel talker and implicit knowledge groups may have been more similar to one another during the test phase than they would otherwise have been. This may have reduced any actual effects of implicit talker knowledge. Perhaps even small amounts of exposure to a voice are sufficient to generate familiarity, and large amounts of additional exposure has little added benefit. Using a different testing method might help to address this issue.

The idea that it is explicit rather than implicit knowledge of a talker's voice that most influences speech segregation, as suggested by the experiments presented here, has several implications for theories of speech perception. To date, there have been two primary theories of voice familiarity effects, which have been viewed as being diametrically opposed. One explanation, proposed by Goldinger (1996), is that talker information is stored within the lexicon, as part of episodic traces of words. According to this approach, hearing a known talker will result in the implicit priming of other tokens of speech by that talker, leading to improved recognition. Although Goldinger's theory would limit such advantages to words previously heard in the known voice, Yonan and Sommers (2000) extended this implicit approach to include cases in which there was voice overlap without phonological overlap. The second explanation for voice familiarity effects, articulated by Nusbaum and Morin (1992), is that listeners use stored information about a talker to better calibrate the perceptual system for the vocal characteristics of the talker. This approach suggests that talker benefits should be present irrespective of whether the particular words had been heard previously; however, it also implies that talkers might need to recognize which voice they were hearing before being able to make use of voice-specific information. That is, calibration depends on identifying the particular talker (or set of reference information). Thus, this approach suggests that there should be effects of explicit talker identification, but not necessarily of implicit familiarity, whereas the previous (implicit) approach predicts exactly the opposite. The present findings suggest the possibility that listeners actually use *both* processes, and that neither theory provides a complete account.[2]

Even assuming that explicit information about talker identity results in activation of a stored representation for that voice, it is not clear how such information would aid in selective attention, or even what information is provided. Presumably, activation of reference information for how an individual speaks could aid in perceptual identification—but how such information would benefit selective attention is less clear. Knowing

[2]Nusbaum and Morin also propose a third alternative, that speech is self-normalizing; according to this approach, acoustic information from the talker can be computed at the time of production, rather than having been prestored. However, this form of recognition is slow and effortful, and as such is only performed when there is frequent variation in talker identity. Since the present study used a single talker for the entire passage, this latter form of talker adjustment is not relevant here.

aspects of a talker's typical speech style (dialectical properties, typical pitch range, etc.) could potentially provide the listener with low-level acoustic cues that could be used to filter the incoming speech signal. However, this presumes that there are obvious acoustic differences between the talker and the background noise; if not, filtering would not serve to isolate the appropriate voice. Future work will need to examine the process by which information about a talker is used in selective attention.

In conclusion, then, this preliminary study suggests that knowledge about a talker appears to be able to influence speech perception abilities in two ways. First, listeners appear to be better able to follow and understand a voice when they know whose voice it is. That is, explicit knowledge of whom they should be attending appeared to help listeners maintain attention on the appropriate signal and interpret it appropriately. In contrast, simply being familiar with a voice did not have this benefit. Second, familiarity with a voice did appear to improve identification accuracy. This effect was present even when the testing situation (a laboratory experiment) was far removed from the situation in which the voice was actually learned. Finding separable effects of talker identity and talker familiarity suggests that there may be more than one manner in which talker-specific information influences perception. Most importantly, however, the present paper serves as a call for considering the different ways that knowledge of a talker's voice can effect both laboratory and real-world performance.

Acknowledgments

The authors thank Dr. Bob Baron for recording the test passages for Experiment 1, Tyler Wunnenberg for recording the distracter passages for Experiment 1, and Ben Schnoor for recording the test passages for Experiment 2. We also thank Sheryl Clouse, Rebecca Ribar, Ben Schnoor, Sarah Stilwill, Laura Veltman, Tyler Wunnenberg and especially Tammy Weppelman for assistance in subject running and coding, and thank Christine Beagle, Robin Nicoletti, and Eva Derecskei for recoding minute-by-minute. We also thank Rachel Smith, Gerry Docherty, and an anonymous reviewer for many helpful comments. Portions of this work were performed when the first author was at the University of Iowa, and was supported by a developmental assignment from the University of Iowa to the first author, and by NIH grant HD37822-01 to the University of Iowa. R. Newman is now at the Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, e-mail: rnewman@hesp.umd.edu.

References

Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language & Speech*, *21*(4), 305–318.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, *2*, 225–235.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45–B53.

Barker, J., & Cooke, M. (1999). Is the sine-wave speech cocktail party worth attending? *Speech Communication*, *27*, 159–174.

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(1), 54–67.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics*, *37*, 483–493.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249.

Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, *44*, 51–55.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, *47*, 191–196.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*, 23–36.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustical Society of America*, *112*(6), 2985–2995.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*(3), 521–533.

Corteen, R. S., & Wood, B. (1972). Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, *94*(3), 308–313.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274–284.

Culling, J. F., & Darwin, C. J. (1993). Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0. *Journal of the Acoustical Society of America*, *93*(6), 3454–3467.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception & Performance*, *30*(4), 643–656.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112–1120.

Dannenbring, G. L., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, *24*, 369–376.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, *5*, 322–337.

Dupoux, E., Kouider, S., & Mehler, J. (2003). Lexical access without attention? Explorations using dichotic priming. *Journal of Experimental Psychology: Human Perception & Performance*, *29*(1), 172–184.

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, *12*(2), 105–111.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory & Cognition*, *5*(6), 658–665.

Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, *5*(5), 499–504.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*(5), 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(1), 152–162.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America*, *111*(3), 1340–1348.

Grose, J. H., Hall, J. W. III., & Mendoza, L. (1995). Perceptual organization in a comodulation masking release interference paradigm: Exploring the role of amplitude modulation, frequency modulation, and harmonicity. *Journal of the Acoustical Society of America*, *97*(5, Part 1), 3064–3071.

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognize you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology*, *51A*(1), 179–195.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, *9*(2), 155–184.

Hirsh, I. J. (1950). The relation between localization and intelligibility. *Journal of the Acoustical Society of America*, *22*(2), 196–200.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(4), 751–763.

Kakehi, K. (1992). Adaptability to differences between talkers in Japanese monosyllabic perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 135–142). Tokyo: Ohmsha.

MacKay, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology*, *25*, 22–40.

Mackersie, C. L. (2003). Talker separation and sequential stream segregation in listeners with hearing loss: Patterns associated with talker gender. *Journal of Speech, Language, & Hearing Research*, *46*(4), 912–918.

Mackersie, C. L., Prida, T. L., & Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *Journal of Speech, Language, & Hearing Research*, *44*(1), 19–28.

Magnuson, J. S., Yamada, R. A., & Nusbaum, H. C. (1994a). Are representations used for talker identification available for talker normalization? Paper presented at the *International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan, September 18–22, 1994.

Magnuson, J.S., Yamada, R.A., Nusbaum, H.C. (1994b). Variability in familiar and novel talkers: Effects on mora perception and talker identification. September 1994 meeting of the Acoustical Society of Japan Technical Committee on Psychological and Physiological Acoustics, Kanazawa, Japan, H-94-44, pp. 1–8.

Magnuson, J.S., Yamada, R.A., Nusbaum, H.C. (1995). The effects of familiarity with a voice on speech perception. Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan, pp. 391–392.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*, 379–390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*, 365–378.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 113–134). Tokyo: Ohmsha.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research on Speech Perception Progress Report 10). Bloomington, IN: Indiana University.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *26*(5), 1228–1244.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*(7), 989–1001.

Ostry, D., Moray, N., & Marks, G. (1976). Attention, practice, and semantic targets. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 326–336.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attribution and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*, 309–328.

Pisoni, D. B. (1990). Effects of talker variability on speech perception: Implications for current research and theory. Paper presented at the ICSLP 90, Kobe, Japan, November 18–22, 1990.

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 143–151). Tokyo: Ohmsha Press.

Pollack, I., & Pickett, J. M. (1958). Stereophonic listening and speech intelligibility against voice babble. *Journal of the Acoustical Society of America*, *30*(2), 131–133.

Poulton, E. C. (1953). Two-channel listening. *Journal of Experimental Psychology*, *46*(2), 91–96.

Remez, R. E., Rubin, P. E., Burns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, *101*, 129–156.

Rivenez, M., Darwin, C. J., & Guillaume, A. (2004). Perception of unattended speech. Paper presented at the *ICAD 04—Tenth meeting of the international conference on auditory display*, Sydney, Australia, July 6–9, 2004.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, & Hearing Research*, *40*(2), 453–463.

Schweinberger, S. R., Herholz, A., & Stief, V. (1997). Auditory long-term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology*, *50A*(3), 498–517.

Shackleton, T. M., & Meddis, R. (1992). The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs. *Journal of the Acoustical Society of America*, *91*(6), 3579–3581.

Sheffert, S. M., & Fowler, C. A. (1995). The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language*, *34*, 665–685.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, synthesized, and reversed speech samples. *Journal of Experimental Psychology: Human Perception & Performance*, *28*(6), 1447–1469.

Spieth, W., Curtis, J. F., & Webster, J. C. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America*, *26*(3), 391–396.

Summerfield, Q., & Assman, P. F. (1991). Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony. *Journal of the Acoustical Society of America*, *89*, 1364–1377.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–248.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics*, *13*, 19–38.

Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics*, *13*, 39–52.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260.

Wood, N. L., Stadler, M. A., & Cowan, N. (1997). Is there implicit memory without attention? A reexamination of task demands in Eich's (1984) procedure. *Memory & Cognition*, *25*(6), 772–779.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, *15*(1), 88–99.