Journal of the American Academy of Audiology

The advantage of knowing the talker

Journal:	*Journal of the American Academy of Audiology*
Manuscript ID:	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Souza, Pamela; Northwestern University, Communication Sciences and Disorders Gehani, Namita; Northwestern University, Wright, Richard; University of Washington, McCloy, Daniel; University of Washington,
Keywords:	Hearing aids and assistive listening devices, Speech perception, Aging, Familiarity

The advantage of knowing the talker

Pamela Souza and Namita Gehani

Roxelyn and Richard Pepper Department of Communication Sciences and Disorders,

Northwestern University, Evanston, IL

Richard Wright and Daniel McCloy

Department of Linguistics, University of Washington, Seattle, WA

Submitted: August 8, 2012

Contact information: Pamela Souza, PhD, Dept. of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208

Email: p-souza@northwestern.edu. Phone: 847-491-2433. Fax:

Work presented in this paper was supported by NIH grants R01DC60014 and R01DC012289.

This data has not been presented previously.

Key words: aging; hearing loss; familiarity; speech recognition; benefit

Abbreviations: dB SNR = decibel signal-to-noise ratio; dB HL = decibels hearing level; ANOVA = analysis of variance

Background: Every audiologist has observed a situation where a patient seems to understand something spoken by his/her spouse but not the same information spoken by a stranger. However, it is not clear whether this observation reflects choice of communication strategy or a true benefit derived from the talker's voice.

Purpose: The current study measured the benefits of long-term talker familiarity for older individuals with hearing impairment in a variety of listening situations.

Research Design: In Experiment 1, we measured speech recognition with familiar and unfamiliar voices when the difficulty level was manipulated by varying levels of a speech-shaped background noise. In Experiment 2, we measured the benefit of a familiar voice when the background noise was other speech (informational masking).

Study Sample: A group of 31 older listeners with high-frequency sensorineural hearing loss participated in the study. Fifteen of the participants served as talkers, and sixteen as listeners. In each case, the talker-listener pair for the familiar condition represented a close, long-term relationship (spouse or close friend).

Data Collection and Analysis: Speech-recognition scores were compared using controlled stimuli (low-context sentences) recorded by the study talkers. The sentences were presented in quiet and in two levels of speech-spectrum noise (Experiment 1) as well as in multitalker babble (Experiment 2). Repeated-measures analysis of variance was used to compare performance between the familiar and unfamiliar talkers, within and across conditions.

Results: Listeners performed better when speech was produced by a talker familiar to them, whether that talker was in a quiet or noisy environment. The advantage of the familiar talker was greater in a more adverse listening situation (i.e., in the highest level of background noise), but was similar for speech-spectrum noise and multi-talker babble.

Conclusions: The present data support a frequent clinical observation: listeners can understand their spouse better than a stranger. This effect was present for all our participants and occurred under strictly controlled conditions in which the only possible cue was the voice itself, rather than under normal communicative conditions where listener accommodation strategies on the part of the talker may confound the measurable benefit. The magnitude of the effect was larger than shown for familiar voices in previous work, suggesting that older listeners with hearing loss who inherently operate under deficient auditory conditions can benefit from a rich communication environment over many years of a relationship.

The advantage of knowing the talker

Every audiologist has observed a situation where a patient seems to understand something spoken by his/her spouse but not the same information spoken by a stranger. We know from both anecdotal observations and research (Scarinci et al., 2008) that spouses or family members make accommodations for the hearing-impaired conversation partner. In long-term relationships in which one partner has hearing impairment, the talker may raise her voice, speak more clearly, or choose vocabulary or syntax that the listener is familiar with. We also know that such strategies have been shown to improve communication within the couple (Preminger, 2008). In other words, we may be observing communication strategies rather than voice familiarity *per se*.

However, there is also evidence that familiarity with a talker's voice can improve speech recognition under some circumstances. Research in this area has focused on deliberately-trained recognition of previously unknown speakers, usually by young listeners with normal hearing. In general, those data indicate that speech recognition is improved when stimuli are produced by a previously-heard talker compared to a novel talker (Palmeri et al., 1993; Nygaard and Pisoni, 1998; Yonan and Sommers, 2000; Sheffert et al., 2002; Newman and Evers, 2007). To understand why this happens, consider the wide variability in speech production. Listeners are able to identify speech sounds despite intra-talker differences in fundamental frequency, speech rate, frequency location of vocal tract resonances, and overriding voice qualities such as tremor or hoarseness. Originally, this information was viewed as a type of noise that listeners had to discard in order to abstract the identifiable components of speech for comparison to stored lexical representations. More recent work has argued that information about the talker is stored and may be used in combination with lexical information (Palmieri et al., 1993; Nygaard et al., 1994, 1995; Goldinger, 1996; Remez et al, 2007). This second view is consistent with the idea that a familiar voice might aid in speech recognition.

The extent to which a voice is familiar will depend on the duration and circumstances of exposure. Because of practical constraints, most published studies evaluated deliberate short-term exposure (hours to days), usually with strictly controlled test material. In those cases, the listener did not know the talker or interact with them in person, but simply heard their recorded voice. In such studies it has been found that mere exposure to a voice does not necessarily confer an advantage. As one example of this, exposure to sentences by a target talker improved recognition of single words by that talker, but exposure to single words did not improve recognition of sentences (Nygaard and Pisoni, 1998; Yonan and Sommers, 2000). A plausible explanation for such findings is the idea that familiarity might be aided by exposure to a richer variety of utterances which cover a range of topics, prosody and context, as would occur in a long-term relationship. Indeed, one study which failed to show an advantage of implicit familiarity exploited a short-term relationship (professor/student) which existed under constrained communication conditions (classroom teaching) (Newman and Evers, 2007). Thus, a real-life paradigm of interest to clinicians is the benefit of long-term, implicit learning over many years of interaction within a relationship.

The case of long-term relationships in which one individual has a hearing loss presents an interesting scenario. First, we focus here on *older* talker-listener pairs. There has been less attention to the advantage conferred by a familiar (or trained) voice for older listeners, although the available data hint that older listeners may respond differently. Several studies have identified age deficits for speaker identification (Yonan and Sommers, 2000; Helfer and Freyman, 2008; Rossi-Katz and Arehart, 2009) or auditory priming tasks (Schacter et al., 1994;

Huang et al., 2010). However, Yonan and Sommers (2000) found that older listeners trained over a two-day period derived as much or more benefit from talker familiarity than younger listeners. Taken together, these data suggest that older listeners may process and/or store voice information differently than younger listeners, but that does not preclude using voice information to improve perception. A final point with regard to age is that all of the training studies conducted thus far have used test materials created within the laboratory. Although talker age was usually not specified, we can assume that the materials were produced by "ideal" (and probably younger) talkers. In the real-life situation we consider, the talker's voice may also be subject to age-related changes, such as vocal tremor or hoarseness (Kendall, 2007) which could make the voice more immediately identifiable; or, alternatively, less intelligible.

Second, we consider a situation in which the listener has hearing loss. There is some evidence that the advantage of talker familiarity may be more beneficial under degraded listening conditions. McLennan and Luce (2005) have argued that familiarity effects are more likely to be observed when processing of the input signal is slowed, as might be the case when the signal is impoverished. Available data for normal-hearing younger listeners presented with low-pass filtered speech (Church and Schacter, 1994) or speech in noise (Nygaard and Pisoni, 1998; Yonan and Sommers, 2000) support this idea. Previous work, including those studies focused on older listeners (Yonan & Sommers, 2000; Huang et al., 2010), tested adults with normal- to near-normal auditory thresholds. Thus, available data may underestimate the effects that occur in the older hearing-impaired population.

A major deficit which accompanies most hearing loss in older adults is difficulty understanding speech in noise. This is most apparent in situations where there is both energetic and informational masking (as when the background consists of other speech). Under such circumstances, Newman and Evers (2007) proposed that familiarity with the target voice could aid auditory stream segregation. It is not known what specific aspects of the familiar voice might contribute to this, but we know that listeners can use a wide variety of acoustic cues to familiarity, including fundamental frequency (Church and Schacter, 1994) and formant trajectories (Sheffert et al., 2002). Familiarity aside, we know that ability to track such frequency variations aids source segregation (Binns and Culling, 2007; Miller et al., 2010). Finally, the ability to follow frequency variations is impaired in some older listeners (Souza et al., 2011). Taken together, these data suggest that even if familiarity enhances those cues for younger listeners, that ability may not be the same in older listeners. Because previous work used broad-band noise maskers rather than speech (Yonan and Sommers, 2000), this issue has not been investigated.

The current study addressed the issue of long-term talker familiarity for older individuals with hearing impairment in a variety of listening situations. In each case, the talker-listener pair for the *familiar* condition represented a close, long-term relationship (spouse or close friend). In Experiment 1, we measured the benefit of a familiar voice when the difficulty level was manipulated by controlling the level of a speech-shaped background noise. In Experiment 2, we measured the benefit of a familiar voice when the background noise was other speech (informational masking).

Experiment 1

Participants

Sixteen adult subjects with hearing loss (9 female, 7 male; mean age 71.9 years) were designated as study listeners. All listeners had bilateral sensorineural loss defined as no air-bone gaps greater than 10 dB and normal tympanograms (Wiley et al., 1987; 1996). Listeners had sloping losses, with high-frequency thresholds in the moderate-to-severe range. In all cases, hearing loss was symmetrical and one ear was randomly selected for testing. Individual test-ear audiograms are shown in Table 1. All listeners had normal short-term memory capacity and orientation to time and place as indicated by a score of 28 or better on the Mini-Mental State Exam (Folstein et al., 1975). Each listener also completed speech-in-noise testing using two lists of the QuickSIN administered via earphones with presentation levels and scoring according to published guidelines (Killion et al., 2004).

The criterion for the familiar talker for each listener was to be a frequent communication partner (see Table 1 for details). Seven of the talker-listener pairs were spouses. The remaining talker-listener pairs were long-term friends, where the pair reported communicating at least three hours each week. Because a primary goal of this study was to investigate communication in older listeners and also to maintain consistency across the talker set, the talkers were all older females (mean age 70.2 years). No talker reported any history of voice, speech or language disorders.

All of the talkers and listeners were native speakers of American English. To control for effects of regional dialect (Wright et al., 2006), all of the talkers and listeners had lived in the greater Chicago area for at least 30 years. All individuals participating in the study completed an informed consent process approved by the Northwestern Institutional Review Board, and were compensated at an hourly rate for their time.

Material

Sentences were low-context sentences drawn from the IEEE corpus. A set of 200 sentences were chosen for recording based on avoidance of alliteration or rhyming patterns (e.g., "it's easy to tell the depth of a well"); avoidance of highly marked locutions (e.g., "the juice of lemons makes fine punch"); and a general preference against sentences that had a natural focus or contrast reading, so that all sentences would be read with similar declarative prosody. The recording apparatus consisted of a head-worn close-talking directional microphone (Shure SM10A) coupled via a microphone amplifier (Rane MS 1S) to an analog-to-digital processor (TDT RX6). Sentences were recorded at a 44.1 kHz sampling rate and quantized at 16 bits. The sentences were read in three randomizations to control for list effects. Each talker was instructed to read the sentences at a natural pace and vocal intensity, without any extra effort to speak clearly or loudly and without giving extra emphasis to any specific words (i.e., to use normal declarative prosody). Vocal level was monitored via a VU meter to ensure sufficient output levels without clipping.

Within each talker, the "best" token of each sentence was chosen as the one absent of any digital clipping or microphone overloading, with a secondary preference for fluent reading (e.g., no unnatural pauses, prosodic consistency) and general clarity, as judged by two trained phoneticians. Each sentence was placed into a separate file, taking care to include the onsets of phonemes with gradual increases in amplitude (such as fricatives) and with low amplitude (such as /h/ and /f/). Each file was padded with 50 ms of silence at the beginning and end of each sentence. Root-mean square levels were normalized across the entire sentence set (200 sentences x 16 talkers).

Procedure

For Experiment 1, 120 of the 200 recorded sentences were selected for testing. For each listener, the sentences were randomly assigned across five talkers and three conditions. The five talkers included the talker familiar to that listener plus four unfamiliar talkers, with the additional four talkers randomly selected from the remaining set of unfamiliar talkers. Listeners were not explicitly told that their familiar talker would be part of the set. The test conditions included sentences in quiet and in two levels of speech-shaped noise (+2 and +6 dB signal-to-noise ratio [SNR]). The noise was a broad-band noise shaped to match the long-term spectrum of all sentences x all talkers, generated using locally-developed Matlab code.

Listeners were seated in a double-walled sound-treated booth. Stimuli were presented over an Etymotic Research ER-2 insert headphone in the test ear. In each trial, the sentence was presented at the specified presentation level dB SPL (dB RMS in a 2cc coupler) and the noise level was adjusted to the desired SNR. Sentences were nominally presented at 35 dB above the listener's pure-tone average at .5, 1 and 2 kHz. First, the listener was familiarized with the task by listening to and repeating 20 different IEEE sentences presented in the same conditions but spoken by a different set of talkers. If during the familiarization procedure a listener indicated that 35 dB SL was uncomfortable, the level was reduced. Across all listeners, the minimum presentation level was 30 dB SL re: pure-tone average. No frequency shaping was used. Listeners were instructed to repeat the sentences, stating individual words recognized. No feedback was provided. Sentence scoring of the five key words per sentence was performed by the tester seated outside the sound booth. Words were counted as correct if spoken in any order, and appended grammatical morphemes were disregarded.

Results

Figure 1 illustrates the effect of familiarity and noise level on recognition accuracy (percent correct). Listeners performed better when speech was produced by a talker familiar to them, even for speech in quiet. The advantage of familiarity was substantial—as much as 15%, on average, depending on the test condition.

Data were analyzed using a two-way (familiarity x SNR) repeated measures analysis of variance (ANOVA)[1]. As expected, overall performance decreased with increasing noise level ($F_{2,156}$=341.16, p<.005). Post-hoc analysis using paired t-tests indicated significantly different scores between quiet and +6 dB SNR, and between +6 dB and +2 dB SNR (p<.005 in each case). Across all conditions, listeners had more difficulty recognizing sentences spoken by unfamiliar talkers ($F_{1,78}$)=22.20, p<.005. The interaction was not significant ($F_{2,156}$=.83, p=.411), suggesting that when considered in terms of absolute performance, the benefit of familiarity was similar across noise conditions.

One weakness of the above analysis is that the variability in absolute intelligibility across listeners exceeded the within-listener effect of familiarity. To understand this issue, consider Figure 2, which shows individual scores for each listener. Each data "column" across the abscissa represents scores for a single listener, where each symbol shows scores for a different talker. Every listener obtained the best performance with their familiar talker. This occurred in all noise conditions. The magnitude of that difference varied from listener to listener. Most relevant to the ANOVA results described above, note that some listeners are simply better speech recognizers than others. This variability among older listeners, even those with similar audiograms, is well-established in the literature. Reporting an absolute-score analysis (i.e., comparing the vertical distribution of all filled triangles to the vertical distribution of all open circles) introduces considerable listener-to-listener variability which may disenfranchise the familiar-to-unfamiliar differences. Accordingly, we calculated the *familiarity benefit* for each listener by calculating the difference in performance between their familiar talker and each unfamiliar talker they heard. The effect of familiarity benefit in the three noise conditions is

[1] For all statistical analyses, raw data were transformed to rationalized arcsine units (RAUs; Studebaker, 1985) to normalize variance across the score range. In cases where the assumption of homogeneity was violated, the values reported follow the Greenhouse-Geisser correction.

shown in Figure 3. A one-way ANOVA indicated a significant effect of SNR ($F_{2,126}$=6.53, p=.002). The benefit of familiarity was greater at +2 dB SNR than at +6 SNR (t_{63}=3.71, p<.005). The benefit of the familiar talker was statistically equivalent in quiet and at +6 dB SNR (t_{63}=-.07, p=.943). Note, however, that this pattern may have been influenced by a ceiling effect for the quiet condition, and perhaps to a lesser extent for the +2 dB SNR condition.

Discussion

The perceptual advantage gained from familiarity with the talker was substantial. One marker of this advantage was that scores for sentences spoken by a *familiar* talker in a +2 dB SNR were nearly identical to scores for an *unfamiliar* talker in a +6 dB SNR—a 4 dB improvement in performance (see Figure 1). Although we expected to see some benefit of familiarity, the magnitude was impressive considering the strict controls to test materials. Indeed, every listener obtained this benefit to some extent: for every listener, scores for their familiar talker (filled triangles in Figure 2) were higher than scores for each of the unfamiliar talkers they heard (open circles in Figure 2). Moreover, since the talkers were reading sentences in a recording booth, rather than talking directly to their spouses or friends, the benefit seen cannot be due to partner-specific conversational or communicative strategies. In other words, the familiarity benefit—and its enhancement in more challenging conditions—is due to abilities or strategies on the part of the listener, not the talker. The fact that this familiarity benefit was stronger in the most adverse noise condition suggests that talker familiarity benefits are robust in noise; or equivalently, the advantage conferred by familiarity can be triggered and exploited even by very brief glimpses of the target signal that "pop out" above the masker noise. To explore this further, Experiment 2 used a multitalker babble noise that represented an everyday listening environment. Babble fluctuates in both the spectral and temporal domains, so may afford greater opportunities for glimpsing. However, it also introduces an informational masking component whereby the background is comprised of meaningful (although not intelligible) speech and which may create a more adverse listening situation for these older listeners with hearing loss (e.g., Pichora-Fuller and Souza, 2003). Based on the finding that the benefit of familiarity was lower in less adverse listening conditions (quiet and +2 dB SNR), we expected that increased task difficulty, if it occurred, might alter the extent of familiarity benefit.

Experiment 2

Participants. Thirteen of the 16 listeners from Experiment 1 (excluding listeners 2, 11 and 13, who were not available) participated in Experiment 2.

Materials. Test materials were the 80 recorded sentences not previously used in Experiment 1. Each sentence was presented in either speech-shaped noise or multitalker babble, both at a +2 dB SNR. The speech-shaped noise was the same noise as used in Experiment 1. Note that the speech-shaped noise condition for Experiment 2 replicated one of the test conditions in Experiment 1, albeit with a different set of unfamiliar talkers for each listener, and served as a reliability check of the Experiment 1 data. The multi-talker babble was a six-talker babble taken from the recording of the Connected Speech Test (Cox et al., 1987), spectrally shaped to be identical to the long-term average spectrum of the sentence set and of the speech-shaped noise. As in Experiment 1, each listener heard sentences spoken by a total of five talkers (the talker familiar to them, plus four unfamiliar talkers). The unfamiliar talkers were chosen randomly and did not repeat any of the unfamiliar talkers heard by that listener in Experiment 1. Each listener therefore heard a total of eight test sentences (40 key words) per talker x noise type. Sentences were blocked by noise type. As in Experiment 1, a set of 20 practice sentences from a different set of talkers was presented to familiarize the listener with the task; half of the practice sentences were presented with speech-shaped noise and half with multitalker babble. Sentence presentation and scoring were otherwise identical to Experiment 1.

Results

Figure 4 shows results, grouped by noise type. Overall performance was higher in speech-spectrum noise than in multitalker babble. For both noises, scores were approximately 15% better for sentences spoken by a familiar talker. This was confirmed with a two-way repeated-measures ANOVA (noise type x familiarity): there was a significant effect of noise type ($F_{1,63}$=216.54, p<.005) and of familiarity ($F_{1,63}$=14.76, p<.005). The difference was similar for both noise types ($F_{1,63}$=.38, p=.537).

Figure 5 shows individual scores for each talker, plotted as in Figure 2. As for Experiment 1, performance was always best with the listener's familiar talker. In general, the pattern of performance is quite similar across noise types, except that performance is lower for the multitalker babble. However, we can also see that some listeners are more susceptible to the babble vs. speech spectrum noise difference than other listeners. For example, compare the effect of babble on listeners 10 and 204; the latter shows considerably less effect of the babble. This may reflect differences in susceptibility to informational masking; "glimpsing" ability; top-down processing; or other mechanisms.

Discussion

Contrary to our hypothesis, we did not find that the group benefit of familiarity was greater when another layer of complexity—informational masking—was added by changing to a multitalker babble; at least at the +2 dB SNR used here. Multitalker babble should increase informational masking, but may have reduced energetic masking due to the opportunities for spectral and temporal glimpsing that occur at a greater rate than for speech-spectrum noise

(Rosen et al., submitted). It is possible that the offset of informational and energetic masking combined to create a condition of similar difficulty. However, this seems unlikely since overall scores were much lower in the multitalker babble. From these data, it's more likely that subjects had simply reached the maximum advantage that could be gained from familiarity.

General discussion

The present data support a frequent clinical observation: listeners can understand their spouse better than a stranger. This effect was noted in all our participants, regardless of their degree of hearing loss and whether the relationship was with a spouse or close friend. The effect persisted even when measured under strictly controlled conditions in which the only possible cue was the voice itself, rather than under normal communicative conditions where listener accommodation strategies on the part of the talker may confound the measurable benefit. And, the effect was implicit: although participants were aware that a close friend or family member had been involved in a previous study visit, they were told only that they would hear sentences spoken by a variety of different speakers, and were not told that they would hear a recognizable voice. Indeed, upon being debriefed, some listeners reported that they had not recognized any of the voices they heard.

Our paradigm probably underrepresented the benefit that might occur under normal communications. For one thing, in everyday listening the listener can often or usually see the talker. This both activates explicit familiarity (shown to provide an even greater advantage; Newman and Evers, 2007) and offers multimodal input, with visual cues to familiarity as well as auditory cues. When one person talks for a period of time, rather than interspersing talkers as was done here, the presentation of a single talker is effectively "blocked", and that manipulation should enhance the ability to recognize a familiar voice (Magnuson et al., 1995). With regard to the test materials themselves, we controlled to a degree for linguistic-prosodic variables in that we used declarative sentences. Finally, when materials are not as controlled as they were here, listeners also have the advantage of shared experience.

An interesting comparison can be made to work by Newman and Evers (2007), a study with some similarities to the present study. In their "implicit familiarity" condition, listeners were presented with material produced by unfamiliar talkers or by a talker with whom they had a previous relationship, but were not told who the speaker would be. In that case, the listeners were university students who had attended a course taught by the familiar speaker. In contrast to the present data, Newman and Evers found no benefit of implicit familiarity. Several differences are of interest here. First, Newman and Evers tested younger listeners with normal hearing using a high-context task (narrative shadowing) at a +5 dB SNR. Considering that normal-hearing young adults can recognize even low-context speech at less favorable SNRs, Newman and Evers' paradigm would be construed as relatively easy (i.e., high redundancy). McLennan and Luce (2005) have posed that the benefits of familiarity will be greatest in conditions in which processing is slowed; i.e., low redundancy conditions. That idea is generally supported by the present data, which showed a high degree of benefit from listening to a familiar voice in a situation which combined high-frequency hearing loss and speech in background noise.

Second, we believe that the relationship exploited by Newman and Evers would have involved minimal familiarity. Good experimental controls were used, in that students who reported they seldom attended class were excluded. However, students would have been exposed to the speaker's voice for several hours each week, over a few months. In those communications, the instructor's topic focus, vocabulary and emotion were likely to have been constrained. Moreover, in our experience as college instructors, attending class—particularly the type of introductory psychology class sampled here—does not guarantee a close focus on the speaker's voice. This represents a very different relationship than tested in the present study, where the listener actively communicated with the talker in a rich and varied conversational environment over a long period of time.

Finally, many college lectures are presented in a large reverberant space, sometimes combined with projection of the speaker's voice through a low-fidelity sound system. This raises the possibility that the voice source characteristics could have been slightly different when presented through the high-fidelity recording system used for study materials compared to the listener's previous experience with that voice. Because we know that voice source characteristics are a determinant of familiarity (Sheffert et al., 2002), this represents another possible difference between previous data and the present study.

The source of familiarity

An interesting experimental question is what conveys familiarity? Source characteristics of the talker's voice (such as fundamental frequency and glottal harmonics) have been shown to be important in learning to identify a new talker (Sheffert et al., 2002). Therefore, we think that source characteristics are important to the familiar-talker benefit that we have observed. However, listeners can also identify individual voices on the basis of fine-structure or phonetic properties when voices are processed to remove source characteristics such as fundamental frequency (Remez et al., 1997; Fellowes et al., 1997). Dynamic aspects of the voice such as intonation (Church and Schacter, 1994) and perhaps speech rate (Bradlow and Nygaard, 1996; Bradlow et al., 1999) may also play a role in talker recognition. Indeed, listeners are able to judge similarity of "voices" after sinewave resynthesis, suggesting that global properties such as prosodic rhythm may play a role in talker identification (Remez et al., 2007) and the concomitant familiar-talker benefit..

This question can also be informed by research on signal familiarity that is not limited to a specific talker. For example, listeners perform better at recognizing speech produced in their own regional dialect (Wright et al., 2006). In that case, aspects of familiarity may include phonemic or allophonic segmental differences known to vary across dialects (e.g., pre-velar /æ/ raising (Dahan et al., 2008). Regional/dialect differences may also include prosodic differences such as speech rate or pitch accent placement or direction (Clopper and Smiljanic, 2011).

Similarly, it is known that talker identification training and familiar talker advantage is more successful when framed in a language the listener understands (Winters et al., 2008; Perrachione et al., 2009; 2011; Levi et al., 2011). One explanation for this effect is that word recognition gives listeners a "toehold" on which to base judgments of talker variation (Perrachione et al., 2009; 2011). Our findings relative to talker familiarity might reflect the same process, although in the reverse direction (leveraging talker identification to help with lexical identification, instead of the other way around).

Within-listener variation and the familiarity benefit

There was a clear group benefit of hearing a familiar rather than an unfamiliar talker, whether that talker was in a quiet or noisy environment. The advantage of the familiar talker was greater in a more adverse listening situation (i.e., in the highest level of background noise). As Figures 2 and 5 illustrate, the benefit of familiarity varied across individuals. Nevertheless, there was no systematic patterning of the magnitude of the familiarity benefit with the "best" or "worst" speech recognizers. For example, compare listeners 3 and 5 in the left panel of Figure 2, who had very different overall performance but whose familiar talkers provided about the same improvement relative to the unfamiliar talkers. Similarly, compare listeners 8 and 12, whose performance on unfamiliar talkers was equivalent, but who derived different familiarity benefits. Moreover, the familiarity benefit did not depend on the familiar talker being highly intelligible *per se*, because each talker was more intelligible to the listener for whom they were familiar regardless of their intelligibility as an unfamiliar talker to the remaining listeners.

Although the study was not designed to systematically measure effects of age within this older listener group, there was no immediate indication that the benefit of familiarity depended on age. As an example, listeners 5 and 10 derived a similar (and significant) familiarity benefit, although they differed in age by nearly two decades.

From a clinical perspective, perhaps the most salient point is the benefit of familiarity for listeners with hearing impairment under adverse listening conditions. The benefit of familiarity was largest for this population under the most adverse SNRs. However, individuals vary in their ability to hear in noisy environments. As a first step toward relating our results to real-world situations, we compared each listener's familiarity benefit to their QuickSIN scores (obtained as part of their initial screening). Figure 6 shows familiarity benefit, grouped by QuickSIN performance and collapsed across SNR. There was a trend for larger familiarity benefits for listeners who had poorer QuickSIN scores, although this trend was not statistically significant ($F_{2,61}$=1.12, p=.334). Analysis may have been limited by inadequate power for the secondary analysis. Nonetheless, the trend suggests that our observed pattern where familiarity benefits were stronger in more adverse environments (poorer SNRs) might also hold true when that adversity occurs because of degraded abilities of the listener (poorer speech-in-noise ability). Clinically, it would be useful to understand the factors underlying familiarity benefit so they could be exploited to improve communication.

Conclusion

The present data support a frequent clinical observation: listeners can understand their spouse better than a stranger. This effect was present for all our participants and occurred under strictly controlled conditions in which the only possible cue was the voice itself, rather than under normal communicative conditions where listener accommodation strategies on the part of the talker may confound the measurable benefit. The magnitude of the effect was larger than shown for familiar voices in previous work, suggesting that older listeners with hearing loss can benefit from a rich communication environment over many years of a relationship. Understanding the sources of familiarity for this population, and the extent to which familiarity benefits vary across listening situations, presents an interesting area for future work.

Acknowledgments.

Work was supported by the National Institutes of Health (R01DC60014 and R01DC012289). The authors thank Frederick Gallun for sharing Matlab code for some portions of this study, and August McGrath for her assistance with materials preparation.

References

Binns C, Culling J. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am* 122:1765-1776.

Bradlow A, Nygaard L, Pisoni D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept Psychophys* 61:206–219.

Bradlow A, Toretta G, Pisoni D. (1996). Intelligibility of normal speech I: Global and gine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20:255-272.

Bradlow A,Nygaard L. (1996). The effect of talker rate and amplitude variation on memory representation of spoken words. *J Acoust Soc Am* 99:2588.

Church B, Schacter D. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *J Exp Psychol: Learning, Memory and Cognition* 20:521-533.

Clopper C, Smiljanic R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39:237-245.

Cox R, Alexander G, Gilmore C. (1987). Development of the Connected Speech Test. *Ear Hear* 8:119S-126S.

Dahan D, Drucker S, Scarborough R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representation? *Cognition* 108:710-718.

Decoster W, Debruyne F. (1997). The ageing voice: changes in fundamental frequency, waveform stability and spectrum. *Acta Otorhinolaryngologica Belgica* 51:105-112.

Fellowes J, Remez R, Rubin P. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Percept Psychophys* 59:839-849.

Folstein M, Folstein S, McHugh P. (1975). "Mini-mental state": A practical method for graidng the cognitive state of patients for the clinician. *J Psychiatr Res* 12:189-198.

Goldinger S. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol: Learning, Memory and Cognition* 22:1166-1183.

Helfer K, Freyman R. (2008). Aging and speech-on-speech masking. *Ear Hear* 29:87-98.

Huang Y, Xu L, Wu X, Li L. (2010). The effect of voice cueing on releasing speech from informational masking disappears in older adults. *Ear Hear* 31:579-583.

Kendall K. (2007). Presbyphonia: a review. *Curr Opin Otolaryngol Head Neck Surg* 15:137-140.

Killion M, Niquette P, Gudmundsen G, Revit L, Banerjee S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116:2395-2405.

Levi S, Winters S, Pisoni D. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *J Acoust Soc Am* 130:4053-4062.

Magnuson J, Yamada R, Nusbaum H. (1995). The effects of familiarity with a coive on speech perception. *Proceedings of the Meeting of the Acoustical Society of Japan* (pp. 391-392).

McLennan C, Luce P. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *J Exp Psychol: Learning, Memory and Cognition* 31:306-321.

Miller S, Schlauch R, Watson P. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *J Acoust Soc Am* 128:435-443.

Newman R, Evers S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics* 35:85-103.

Nygaard L, Pisoni D. (1998). Talker specific learning in speech perception. *Percept Psychophys* 60:355-376.

Nygaard L, Sommers M, Pisoni D. (1994). Speech perception as a talker-contingent process. *Psychological Sciences* 5:42-46.

Nygaard L, Sommers M, Pisoni D. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Percept Psychophys* 57:989-1001.

Palmeri T, Goldinger S, Pisoni D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *J Exp Psychol* 19:309-328.

Perrachione T, del Tufo S, Gabrieli, J. (2011). Human voice recognition depends on language ability. *Science* 333:595.

Perrachione T, Pierrehumbert J, Wong P. (2009). Differential neural contributions to native- and foreign-language talker identification. *J Exp Psychol: Hum Percept Perform* 35:1950-1960.

Pichora-Fuller M, Souza P. (2003). Effects of aging on auditory processing of speech. *Int J Audiol* 42 suppl 2:2S11-2S16.

Preminger J. (2008). Should significant others be encouraged to join adult group audiologic rehabilitation classes? *J Am Acad Audiol* 14:545-555.

Remez R, Fellowes J, Nagel D. (2007). On the perception of similarity among talkers. *J Acoust Soc Am* 122:3688-3697.

Remez R, Fellowes J, Rubin P. (1997). Talker identification based on phonetic information. *J Exp Psychol Hum Percept Perform* 23:651-666.

Rosen S, Souza P, Ekelund C, Majeed A. Listening to speech in a background of other talkers: effects of talker number and noise vocoding (submitted).

Rossi-Katz J, Arehart K. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *J Speech Lang Hear Res* 52:435-453.

Scarinci N, Worrall L, Hickson L. (2008). The effect of hearing impairment in older people on the spouse. *Int J Audiol* 47:141-151.

Schacter D, Church B, Osowiecki D. (1994). Auditory priming in elderly adults: impairment of voice-specific implicit memory. *Memory* 2:295-323.

Sheffert S, Pisoni D, Fellowes J, Remez R. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J Exp Psychol* 28:1447-1469.

Souza P, Arehart K, Miller C, Muralimanohar R. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear* 32:75-83.

Wiley T, Cruickshanks K, Nondahl D, Tweed T, Klein R, Klein B. (1996). Tympanometric measures in older adults. *J Am Acad Audiol* 7:260-268.

Wiley T, Oviatt D, Block M. (1987). Acoustic-immittance measures in normal ears. *J Speech Hear Res* 30:161-170.

Winters S, Levi S, Pisoni D. (2008). Identification and discrimination of bilingual talkers across languages. *J Acoust Soc Am* 123:4524-4538.

Wright R, Bor S, Souza P. (2006). Region, gender and vowel quality: a word to the wise hearing scientist. Paper presented at the Acoustical Society of America. Honolulu, HI.

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

Table 1. Participant characteristics

Listener		Listener hearing thresholds (dB HL) for test ear						Talker		
Study ID	Age (years)	.25	.5	1	2	4	8	Relationship	Duration (years)	Age (years)
001	65	30	20	15	40	55	75	Spouse	42	68
002	82	15	20	15	30	35	75	Spouse	58	80
003	78	35	20	20	30	45	65	Friend	30	61
004	72	25	25	35	45	50	90	Friend	35	70
005	79	45	45	50	55	70	75	Friend	8	82
006	70	20	20	30	40	60	60	Spouse	47	67
007	78	20	30	45	50	55	65	Friend	8	82
008	83	35	45	50	55	50	65	Friend	8	84
009	60	15	20	15	40	55	50	Friend	11	73
010	63	10	15	15	50	70	60	Spouse	44	62
011	53	20	25	25	55	45	55	Spouse	23	55
012	80	25	40	45	60	65	85	Friend	30	68
013	66	15	25	20	25	55	60	Friend	40	66
015	74	30	25	20	25	45	75	Spouse	48	71
016	76	35	40	40	35	50	65	Friend	7	64
204	72	10	15	10	30	60	80	Spouse	52	70

Figures

Figure 1. Speech recognition as a function of signal-to-noise ratio. In each case, the background noise was a broad-band noise shaped to the spectrum of the test sentences. Scores for the familiar talkers are shown by the filled boxes and scores for the unfamiliar talkers by the unfilled boxes. The ends of the box indicate the quartile values and the bar indicates the sample median. The whiskers extend to 1.5 times the height of the box or, if no case/row has a value in that range, to the minimum or maximum values. Outliers greater or less than 1.5 times the box height are plotted as open circles.

Figure 2. Speech-recognition scores for individual listeners. Each data point represents a score for a different talker, with the familiar talkers plotted as filled triangles and the unfamiliar talkers plotted as open circles. Each listener heard one familiar and four unfamiliar talkers. For some listeners identical unfamiliar-talker scores are overlaid on the plot. Panels show the three SNR conditions from Experiment 1.

Figure 3. Familiarity benefit, calculated as the difference between the familiar and unfamiliar talker for each listener, as a function of signal-to-noise ratio.

Figure 4. Speech recognition as a function of noise type. In each case, the background noise was presented at +2 dB signal-to-noise ratio. Scores for the familiar talkers are shown by the filled boxes and scores for the unfamiliar talkers by the unfilled boxes.

Figure 5. Speech-recognition scores for individual listeners. Each data point represents a score for a different talker, with the familiar talkers plotted as filled triangles and the unfamiliar talkers plotted as open circles. Each listener heard one familiar and four unfamiliar talkers, but

for some listeners cases identical unfamiliar-talker scores are overlaid on the plot. Panels show the two noise types from Experiment 2.

Figure 6. Mean familiarity benefit (error bars indicate +/- one standard error) from Experiment 1, as a function on listener speech-in-noise ability (measured with QuickSIN).

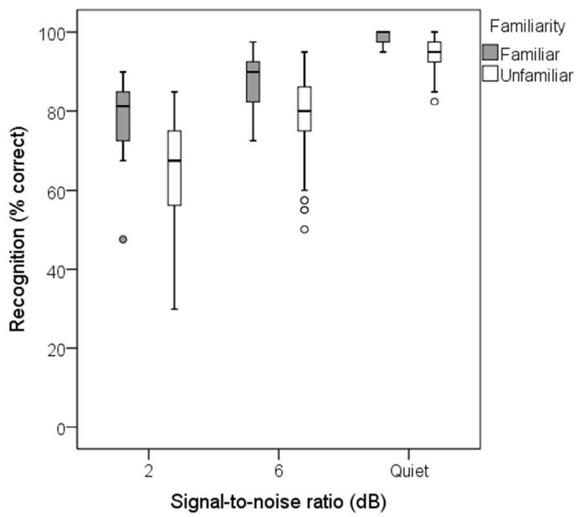

Figure 1

Figure 1
254x190mm (96 x 96 DPI)

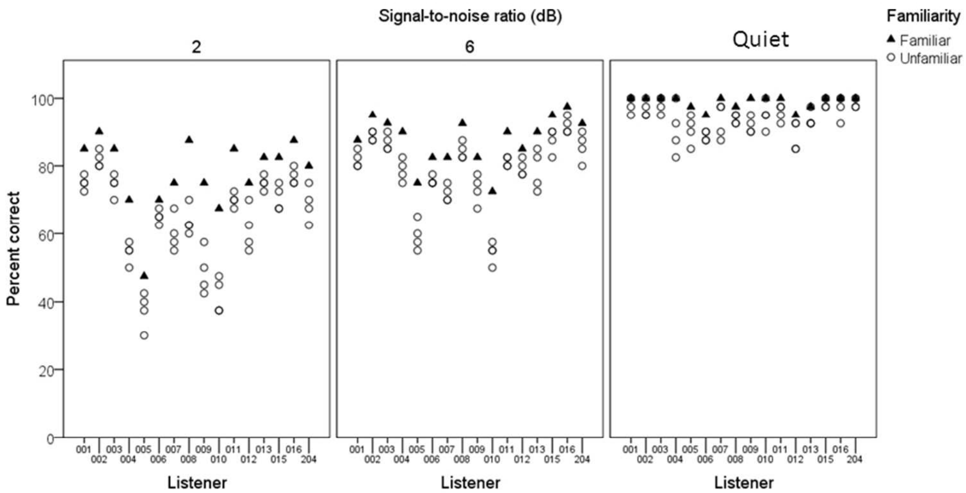

Figure 2

Figure 2
254x190mm (96 x 96 DPI)

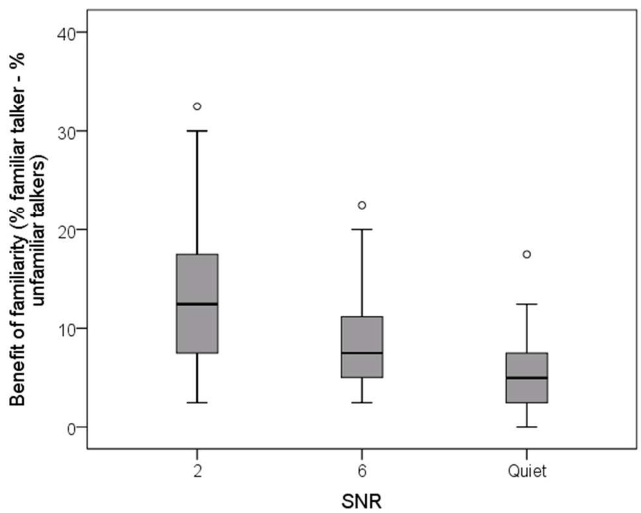

Figure 3

Figure 3
254x190mm (96 x 96 DPI)

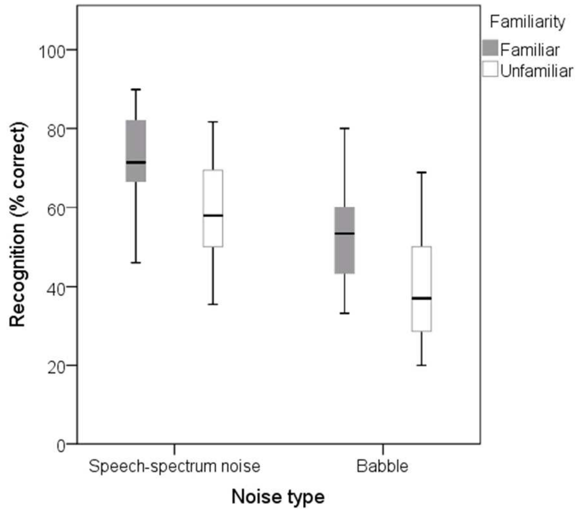

Figure 4

Figure 4
254x190mm (96 x 96 DPI)

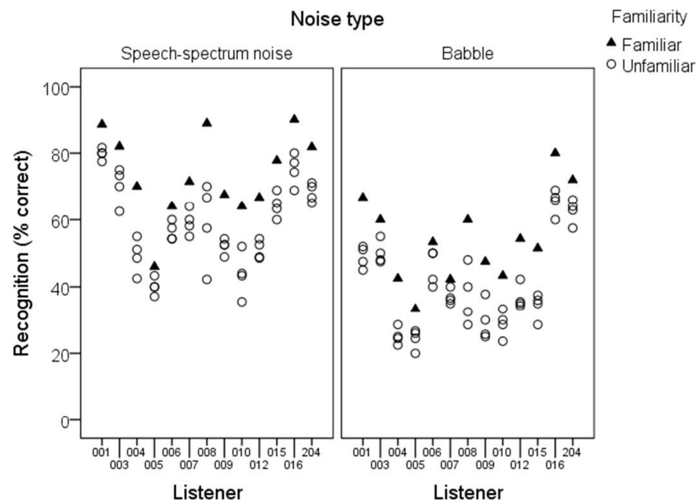

Figure 5

Figure 5
254x190mm (96 x 96 DPI)

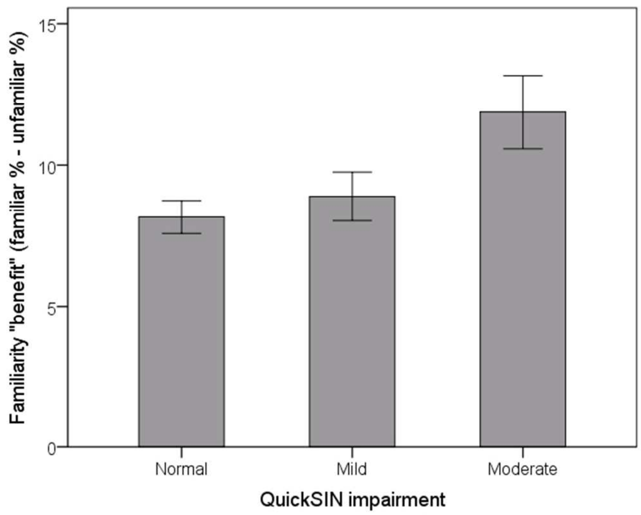

Figure 6

Figure 6
254x190mm (96 x 96 DPI)