Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice

Ingrid S. Johnsrude[1,2], Allison Mackey[1], Hélène Hakyemez[1], Elizabeth Alexander[1], Heather P. Trang[1], and Robert P. Carlyon[3]

[1]Department of Psychology, Queen's University; [2]Linnaeus Centre for Hearing and Deafness, Linköping University; and [3]MRC Cognition & Brain Sciences Unit, Cambridge, England

Abstract

People often have to listen to someone speak in the presence of competing voices. Much is known about the acoustic cues used to overcome this challenge, but almost nothing is known about the utility of cues derived from experience with particular voices—cues that may be particularly important for older people and others with impaired hearing. Here, we use a version of the coordinate-response-measure procedure to show that people can exploit knowledge of a highly familiar voice (their spouse's) not only to track it better in the presence of an interfering stranger's voice, but also, crucially, to ignore it so as to comprehend a stranger's voice more effectively. Although performance declines with increasing age when the target voice is novel, there is no decline when the target voice belongs to the listener's spouse. This finding indicates that older listeners can exploit their familiarity with a speaker's voice to mitigate the effects of sensory and cognitive decline.

Keywords

The listening conditions of everyday life present the auditory system with a formidable challenge—decomposing a complex waveform containing information about multiple sounds so that one source, such as a voice, can be identified, tracked, and understood. This "cocktail-party" problem (Cherry, 1953) presents a significant challenge to young, healthy listeners with normal hearing, and it is particularly problematic for hearing-impaired and older listeners, who typically have much more difficulty understanding speech in the presence of background sound than they do understanding speech in quiet (Gatehouse & Noble, 2004; van Rooij, Plomp, & Orlebeke, 1989).

Much of the research on the cocktail-party problem has focused on identifying the physical, acoustic characteristics (such as pitch, timbre, and location) that listeners can and cannot exploit to separate competing sounds (e.g., Bregman, 1990; Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001; Darwin & Carlyon, 1995; Hawley,

Litovsky, & Culling, 2004; Plomp & Mimpen, 1981). Some influence of experience-based processes has been noted: Listeners benefit when they know the content of the speech that they are trying to identify (Bregman, 1990), when they know the sex of the target talker (Brungart et al., 2001), when masking speech is not in the listener's native language (Cooke, Garcia Lecumberri, & Barker, 2008; van Engen & Bradlow, 2007), and when the target talker is familiar (Barker & Newman, 2004; Magnuson, Yamada, & Nusbaum, 1995; Newman & Evers, 2007; Nygaard & Pisoni, 1998; Yonan & Sommers, 2000). For example, Japanese listeners are more accurate at transcribing moras spoken by family members than by strangers (Magnuson et al., 1995) when those moras,

Corresponding Author:
Ingrid S. Johnsrude, Queen's University, Department of Psychology, 62 Arch St., Kingston, Ontario K7L 3N6, Canada
E-mail: ingrid.johnsrude@queensu.ca

although slightly degraded, are presented in isolation and with no masker.

A benefit based on identifying a message uttered by a familiar voice may be mediated by processes unrelated to sound segregation. Listeners may have an increased tendency to attend to a familiar voice or, in open-set tasks, to guess when unsure of what a familiar, rather than a novel, voice has said. Another possible mechanism, not related to segregation, is that familiarity may enable people listening to a mixture of sounds to extract one that matches a preexisting mental template (Bregman, 1990). Note, however, that none of these alternative explanations predict a benefit from familiarity with the *masker*. Here, we studied the effects of both masker and target familiarity when the listener did not know in advance whether the target, the masker, or neither voice would be familiar. If familiarity with a masker can benefit intelligibility of a novel target voice, this is most probably due to voice knowledge influencing sound segregation itself and facilitating the perceptual separation of voice streams to permit more accurate tracking and perception of one stream. We examined the effect that hearing a familiar voice had on perception by middle-aged and older listeners, whose ability to listen in noisy situations is degraded by peripheral hearing loss but who may be able to compensate by using knowledge of a familiar voice, such as that of their spouse.

We used an adaptation of the coordinate-response-measure (CRM) procedure (Bolia, Nelson, Ericson, & Simpson, 2000; Brungart, 2001; Brungart et al., 2001); see Figure 1. On each trial, the participant heard two concurrent sentences of the form "Ready [call sign] go to [color] [number] now," with each sentence spoken in a different voice. The participant then reported the color and number spoken by the talker who uttered the call sign "Baron" (the target). To select the correct voice, the listener first had to segregate the word "Baron" from the competing call sign and then track that voice until both the color and number uttered by that talker had been identified. The target voice, the masker voice, or neither voice was that of the participant's spouse. By comparing performance when the spouse's voice was the target (with a novel-voice masker) with performance in a "novel-baseline" condition in which both voices were from strangers, we could measure the benefit of the spouse's voice being the focus of attention. Such benefit would be consistent
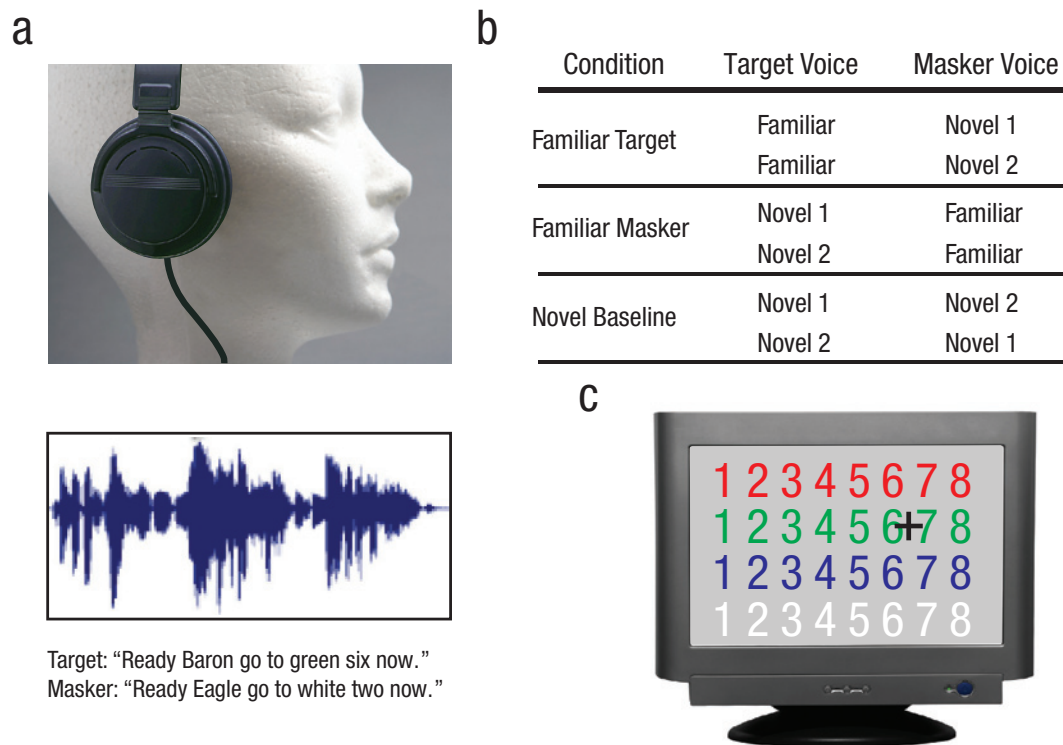
a

Target: "Ready Baron go to green six now."
Masker: "Ready Eagle go to white two now."

b

Condition	Target Voice	Masker Voice
Familiar Target	Familiar	Novel 1
	Familiar	Novel 2
Familiar Masker	Novel 1	Familiar
	Novel 2	Familiar
Novel Baseline	Novel 1	Novel 2
	Novel 2	Novel 1

c

Fig. 1. Procedure used in the listening session. On each trial (a), participants listened diotically over headphones to two concurrently presented sentences spoken in different voices. Participants had to focus on the target sentence, indicated by the call sign "Baron," while ignoring the masker sentence, indicated by one of three other call signs (e.g., "Eagle"). The three conditions in the study (b) varied according to whether the target voice, the masker voice, or neither voice was that of the participant's spouse. The sentences were spoken in either a familiar voice (the participant's spouse) or a novel voice (the spouse of other participants who were sex- and age-matched to the listener). Regardless of condition, the participant tracked the target voice to the color-number coordinate at the end of the utterance (c) and responded by clicking the correctly colored digit on a computer screen.

with reports of better performance on intelligibility tasks for materials spoken in a voice on which listeners have been trained (Newman & Evers, 2007; Nygaard & Pisoni, 1998; Yonan & Sommers, 2000).

Crucially, comparison of performance when the spouse's voice is the masker (and the target voice is novel) with performance in the novel-baseline condition reveals whether listeners can use information about a familiar masking voice to help them track a novel voice. This contrast allows one to rule out explanations for different performance that are related to the nature of the to-be-reported voice, because the target voices in both conditions are novel. Furthermore, the use of a forced-choice task means that any differences between conditions cannot be attributed to an increased tendency to guess when uncertain, as is the case with open-set tasks (Newman & Evers, 2007; Nygaard & Pisoni, 1998).

Method

Participants

Participants were 23 men and 23 women (i.e., 23 married couples) between the ages of 44 and 79 years. To ensure that voices in the familiar-voice conditions would be extremely familiar, we tested couples that had been living together for 18 years or more. The study was cleared by the Queen's University General Research Ethics Board, and informed consent was obtained from all participants after the nature of the study was explained. Given the large age range of participants, we sorted individuals into two groups of approximately equal size: those aged under 60 ($n = 24$) and those 60 and older ($n = 22$). Demographic data for the two groups are given in Table 1.

Materials and procedure

All participants came to the lab for two sessions. The first session involved audiometric screening and stimulus recording. All participants were recorded producing 128 sentences from the CRM database (Bolia et al., 2000). Call signs were "Baron" (the target), "Charlie," "Arrow," and "Eagle"; colors were "red," "green," "blue," and "white"; and numbers were 1 through 8.

Participants returned between a week and a month after the recording session to participate in the listening study, which was usually completed in a single session. Recordings from the first session were used with the same group of participants, this time as listeners. Each listener heard three voices: their spouse's voice and two novel voices. The novel voices were those of other people's spouses, sex- and (approximately) age-matched to the spouse voice (hence, the "swinging" of the article's title), with the limitation that as we recruited and tested in two large "drives," not all voices were available when the first set of couples was tested.

On each trial in the test session, participants heard two different CRM sentences, each of which was spoken simultaneously by a different voice. They were instructed to listen to the phrase with the call sign "Baron" (i.e., the target phrase) and then click on the correct color-number combination identified in the target phrase (see Fig. 1). These color-number combinations appeared on a computer screen in the form of the numbers 1 through 8 ranged sequentially across the screen in four rows. Each row of numbers appeared in one of the four different colors.

Participants were tested in three conditions. In the familiar-target condition, the "Baron" sentence was spoken by the spouse, and one of two possible novel talkers produced the masker sentence (which contained the call sign "Charlie," "Arrow," or "Eagle" and color-number coordinates different from those of the target). The familiar-masker condition involved hearing the target phrase in one of two possible novel voices, with the spouse voice uttering a masker phrase. Each of the two novel voices served an equal number of times in these two conditions. The novel-baseline condition involved hearing the target and masker phrases in the two different novel voices; each novel voice served as target and masker an equal number of times. To ensure that participants would not be able to use the constant amplitude of the target sentence as a cue, we varied the amplitude of the target stimulus across trials while holding the amplitude of the masker stimulus constant. Five target-to-masker ratios (TMRs; −6 dB, −3 dB, 0 dB, +3 dB, and +6 dB) were tested.

Table 1. Demographic Information for the Two Age Groups

Group	n	Sex		Age (years)		Years with spouse	
		Female	Male	M	Range	M	Range
< 60 years	24	14	10	54	44–59	27	18–40
≥ 60 years	22	9	13	67	60–79	35	20–49

Six hundred trials were randomly generated, such that the spouse's voice was the target in 200 trials (familiar-target), the masker in 200 (familiar-masker), and not present in 200 (novel-baseline); see Figure 1b. Trials from these three conditions were equally distributed across TMRs (i.e., 40 trials at each TMR for each condition). Further details regarding the participants, materials, and procedure are given in the Supplemental Material available online.

Data analysis

Trials were considered correct only if both color and number coordinates identified by the participant were from the target. Accuracy was defined as the percentage of trials on which the participant made a correct response. All *F*-test significance values are reported with the Huynh-Feldt correction, and pairwise comparisons are reported with Sidak correction to control Type I error.

Given the broad age range of our participants, the relationship between age and accuracy was also investigated using Pearson correlations, for each condition separately. These were calculated using accuracy scores averaged across TMRs and for individual TMRs.

Errors were categorized into three different types. Errors in which both coordinates came from the masker are called "wrong-voice" errors. Errors in which one coordinate (color or number) came from each of the two presented phrases are called "mixed-voice" errors. The third error type, involving at least one coordinate that was not spoken by either the target or masker voice, comprised only 13% of the errors and was not further analyzed.

Results

As shown in Figure 2, performance was superior in the familiar-target condition, when the spouse's voice was the target (and the masker voice was novel), compared with the novel-baseline condition, in which both voices were novel. This finding provides a measure of the benefit that people obtain when their spouse's voice is the focus of attention. Performance was also better in the familiar-masker condition, when the spouse's voice was the masker (and the target voice was novel), than in the novel-baseline condition, at least when the familiar masker was at least as loud as the target message (i.e., TMR of 0 dB or less).

The trends described above were confirmed by an analysis of variance (ANOVA) with age group and sex as between-subjects factors, and condition and TMR as within-subjects factors. The effect of TMR was significant, $F(2.396, 100.615) = 98.35$, $p < .001$, $\eta_p^2 = .70$: Performance differed significantly among all TMRs ($p < .05$) except for −3 dB versus 0 dB. The effect of condition was also
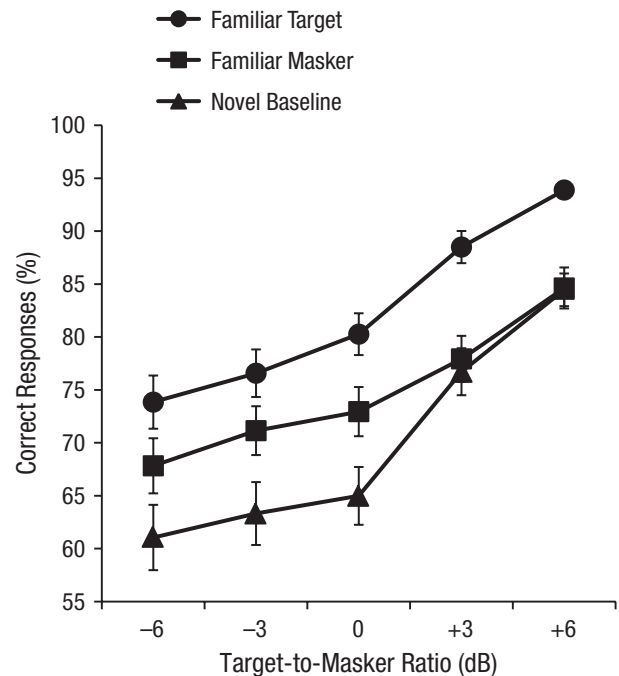

Fig. 2. Percentage of correct responses as a function of target-to-masker ratio and condition. Error bars show standard errors of the mean.

significant, $F(2, 84) = 21.34$, $p < .001$, $\eta_p^2 = .34$, and all three conditions differed significantly from each other (familiar-target vs. familiar-masker, $p < .0025$; familiar-masker vs. baseline, $p < .05$; familiar-target vs. baseline, $p < .001$). There was also a significant interaction between condition and TMR, $F(6.812, 286.099) = 2.98$, $p < .01$, $\eta_p^2 = .07$, attributable to the fact that performance differed among all conditions only at −3 and 0 dB TMR ($p < .05$). In contrast, at TMRs of +3 and +6 dB, performance was very similar in the familiar-masker and novel-baseline conditions, and at a TMR of −6 dB, the difference between the familiar-target and familiar-masker conditions was not significant. The effects of age group and sex, and the interactions involving these factors, were not significant.

Despite the lack of any interaction with age, treated as a categorical variable, we were interested to know whether age was related systematically to performance in any of the conditions. We therefore performed an additional analysis, the results of which showed that the correlation between performance and age depended on condition; see Figure 3. Specifically, performance (averaged across TMR) decreased with increasing age in the novel-baseline condition ($r = −.44$, $p < .005$) and familiar-masker condition ($r = −.35$, $p < .025$), but not in the familiar-target condition ($r = −.032$, n.s.). To compare these correlations, we computed 95% confidence intervals (CIs) for the difference between the correlations
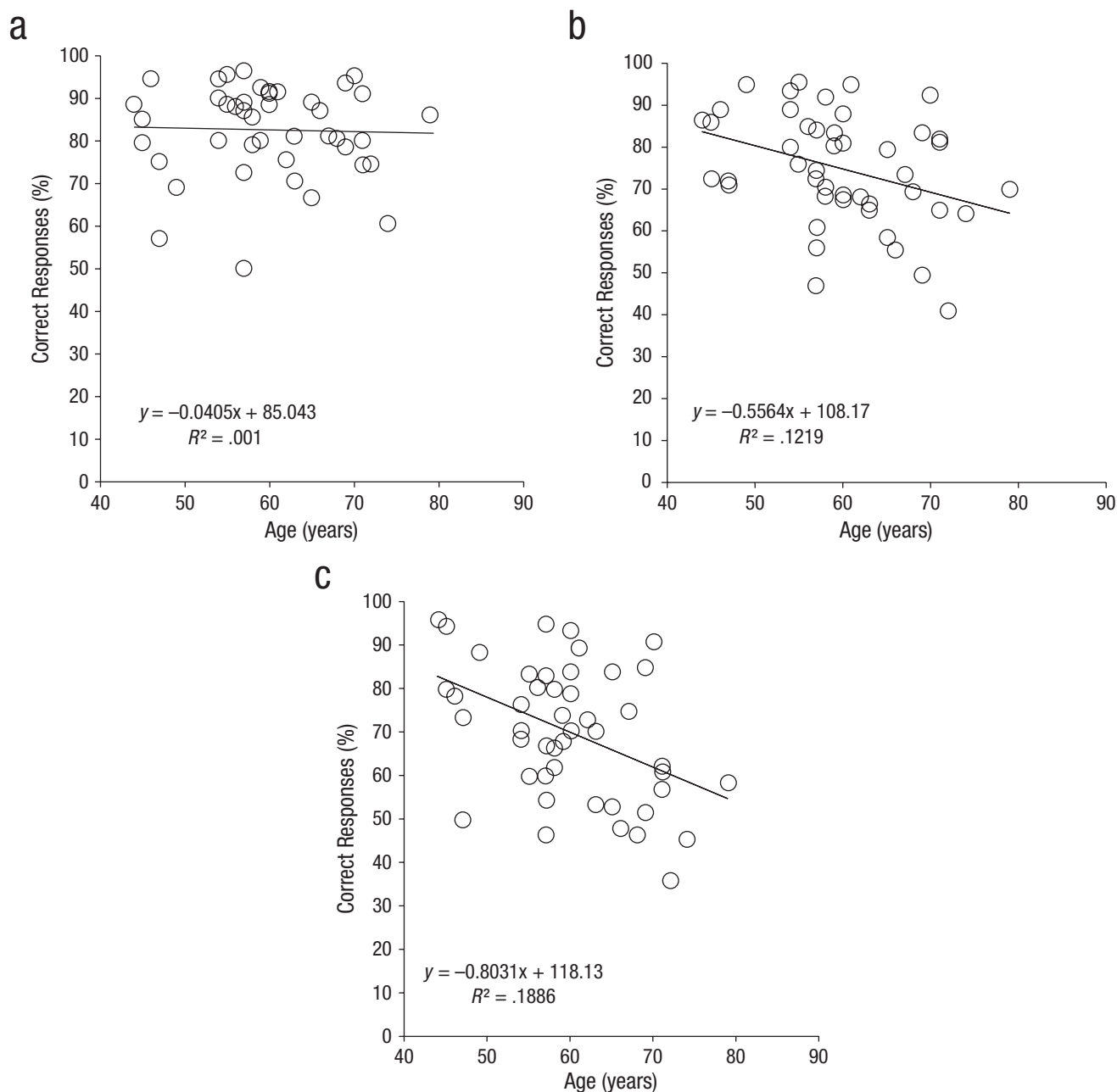
Fig. 3. Scatter plots (with best-fitting regression lines) showing the percentage of correct responses as a function of age, collapsed across target-to-masker ratio. Results are shown separately for the (a) familiar-target, (b) familiar-masker, and (c) novel-baseline conditions.

(Zou, 2007). If the range between the CIs includes 0, then the two correlation coefficients do not differ reliably from each other. According to this measure, only the familiar-target and novel-baseline correlations differed significantly (95% CI = [.10, .68]). The familiar-masker correlation did not differ significantly from the other two.

Although the absence of a significant correlation in the familiar-target condition could have been due to a ceiling effect, this was probably not the case because significant correlations were not obtained in this condition at even

the most disadvantageous TMRs (−3 and −6 dB; Table 2), when performance was clearly below ceiling. Because correlations can be sensitive to the overall level of performance, we also compared correlations between age and performance in two conditions in which performance was approximately equal (at 76.6%): the novel-baseline condition at a TMR of +3 dB and the familiar-target condition at a TMR of −3 dB. There was no correlation between performance and age in the familiar-target condition ($r = .13$, n.s.; Table 2; Fig. 4a), whereas age reliably

Table 2. Correlations Between Age and Percentage of Trials With Correct Responses in the Three Conditions for Each Target-to-Masker Ratio (TMR)

TMR	Novel-baseline condition	Familiar-masker condition	Familiar-target condition
+6 dB	−.28	−.20	−.15
+3 dB	−.43*	−.25	−.11
0 dB	−.41*	−.29*	−.04
−3 dB	−.38*	−.38*	.13
−6 dB	−.43*	−.36*	−.07

*$p < .05$ (two-tailed).

predicted performance in the novel-baseline condition ($r = −.43$, $p < 005$; Fig. 4b). The 95% CI for the difference between these two correlations did not include 0 ([.24, .84]), which indicates that they are reliably different.

The comparison between performance in the familiar-target condition at −3dB and the novel-baseline condition at +3dB is also important because it sheds light on whether the decline in performance with age was simply due to the deterioration in frequency selectivity at the auditory periphery with advancing years (van Rooij et al., 1989) or was instead due to other (cognitive) factors. If frequency selectivity is to blame, then a more marked decline of performance with age should have been evident in the more unfavorable listening conditions of the lower TMR (i.e., −3 dB in the familiar-target condition), in which frequency selectivity was really challenged. In contrast, if the performance decline with age is due to cognitive factors, then it should have been most marked when older listeners could not exploit familiarity, even if the listening conditions were relatively favorable (i.e., +3dB in the novel-baseline condition). The data clearly favor the latter prediction (see Fig. 4). We directly compared performance in these two conditions, within subjects. This showed the degree to which performance was better with a familiar, compared with a novel, target voice (when the masker was novel), equating for overall level of performance. Age significantly predicted this benefit ($r = −.49$, $p < .001$; see Fig. 4c).

The errors that our participants made were overwhelmingly of the wrong-voice type (37% of total errors) or mixed-voice type (50% of total errors). These two error types appear to reflect different processes, as indicated by the different patterns of errors as a function of age group, condition, and TMR; see Figures 5 and 6.

We next performed a univariate ANOVA on the percentage of trials that yielded mixed-voice and wrong-voice errors, including two between-subjects factors (age group and sex) and three within-subjects factors (error type, condition, and TMR). This analysis revealed significant interactions between error type, condition, and TMR, $F(5.666, 237.966) = 3.68$, $p < .005$, $\eta_p^2 = .08$, and between

error type, condition, and age group, $F(2, 84) = 3.15$, $p < .05$, $\eta_p^2 = .07$.

Listeners committed fewer errors of both types in the familiar-target than in the novel-baseline condition across all TMRs (Fig. 5); this finding confirms the robustness of the familiar-target benefit. The familiar-masker benefit was less consistent: Listeners committed fewer wrong-voice errors (Fig. 5b) in the familiar-masker than in the novel-baseline condition, but only when the TMR was less than or equal to 0 dB (ps < .005). No familiar-masker benefit was evident for mixed-voice errors (Fig. 5a) because similar proportions of this type of error were committed in the familiar-masker and novel-baseline conditions across all TMRs.

Approximately equal numbers of mixed-voice errors were committed by listeners in the familiar-masker and novel-baseline conditions, with no effect of age (Fig. 6a). In contrast, there were significantly fewer wrong-voice errors in the familiar-masker condition than in the novel-baseline condition, but only for younger listeners ($p < .005$); see Figure 6b. In other words, the familiar-masker benefit shown in Figure 2 was mainly due to younger listeners making fewer wrong-voice errors than in the novel-baseline condition, whereas older people made similar numbers of both types of error. Compared with older people, younger people were significantly less likely to mistake their spouse's voice than a novel voice for the target, which suggests that they were able to use familiar-voice information even when it was not the focus of attention.

Several other effects were statistically significant. There was a main effect of error type, $F(1, 42) = 21.50$, $p < .001$, $\eta_p^2 = .34$, with more mixed-voice errors ($M = 12.20\%$, $SEM = 0.89$) than wrong-voice errors ($M = 9.20\%$, $SEM = 0.60$). The main effects of TMR and condition were also significant—TMR: $F(2.486, 104.426) = 83.08$, $p < .001$, $\eta_p^2 = .67$; condition: $F(2, 84) = 21.00$, $p < .001$, $\eta_p^2 = .33$—with error rates generally decreasing with increasing TMR and mirroring the accuracy data across conditions (i.e., error rates were highest in the novel-baseline condition, lowest in familiar-target condition, and intermediate in
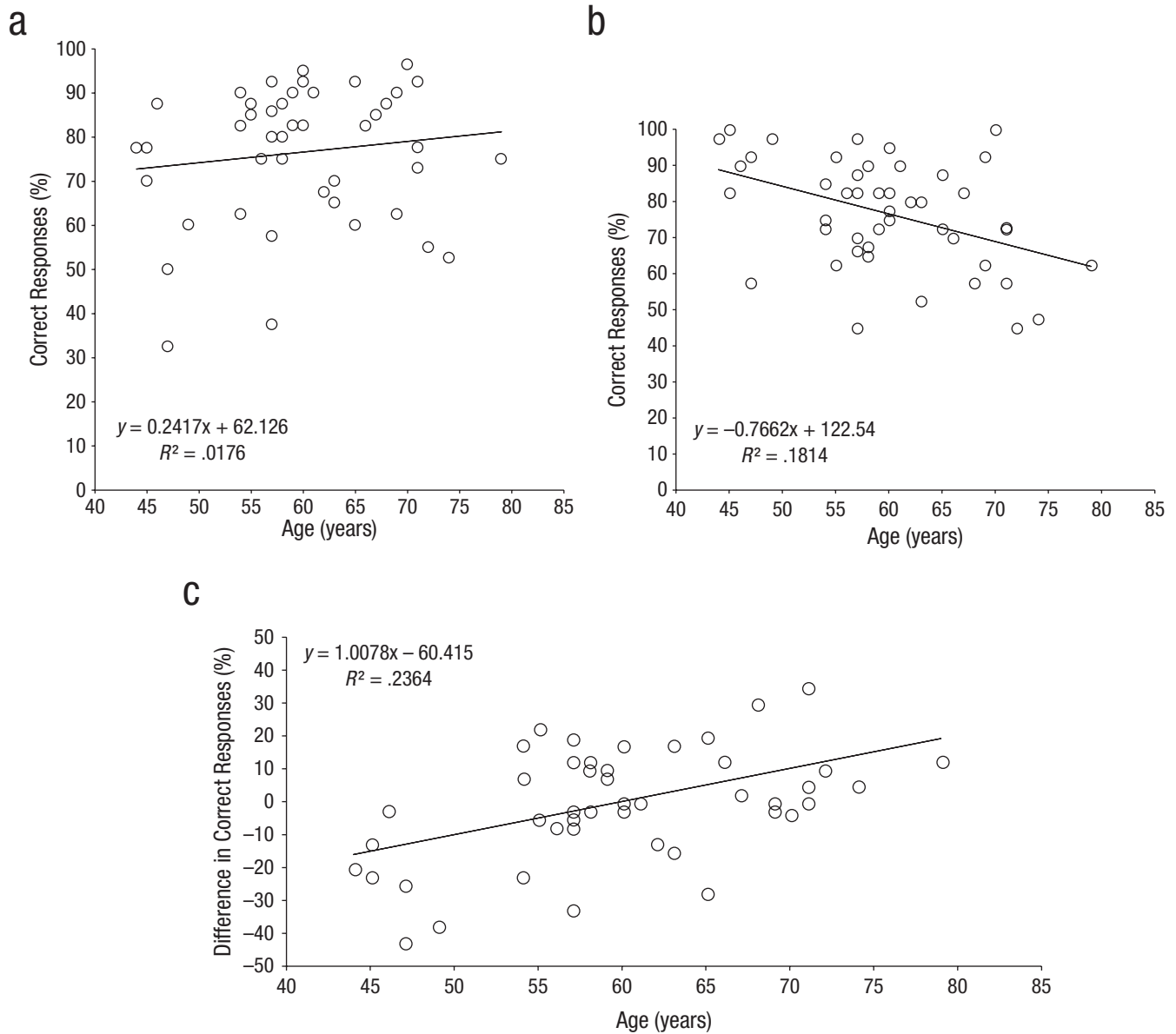
Fig. 4. Scatter plots (with best-fitting regression lines) showing correlations between performance as a function of age in two conditions equated for overall performance. The plots in (a) and (b) show the percentage of correct responses in, respectively, the familiar-target condition at a target-to-masker ratio (TMR) of −3 dB and the novel-baseline condition at a TMR of +3 dB. The plot in (c) shows the difference between the percentage of correct responses in these two conditions.

the familiar-masker condition; all pairwise comparisons significant at $p < .025$). The effects of age group and sex were not significant.

Condition and TMR interacted significantly, $F(6.698, 281.307) = 4.15$, $p < .001$, $\eta_p^2 = .09$, with error rates in the familiar-target and familiar-masker conditions not differing from each other at TMRs of −6 and −3 dB (both different from error rates in the novel-baseline condition, $p < .01$). Error rates in the familiar-masker and novel-baseline conditions did not differ from each other at +3 and +6 dB (both different from familiar-target, $p < .005$);

see Figure 5. The interaction between error type, condition, and sex was also significant, $F(2, 84) = 5.66$, $p < .01$, $\eta_p^2 = .12$. Specifically, men made more wrong-voice (but not mixed-voice) errors than women when the spouse's voice was the target ($p < .025$).

Discussion

Speech perception is considerably improved when a highly familiar voice is present in a two-voice cocktail-party situation. In the present study, a benefit was
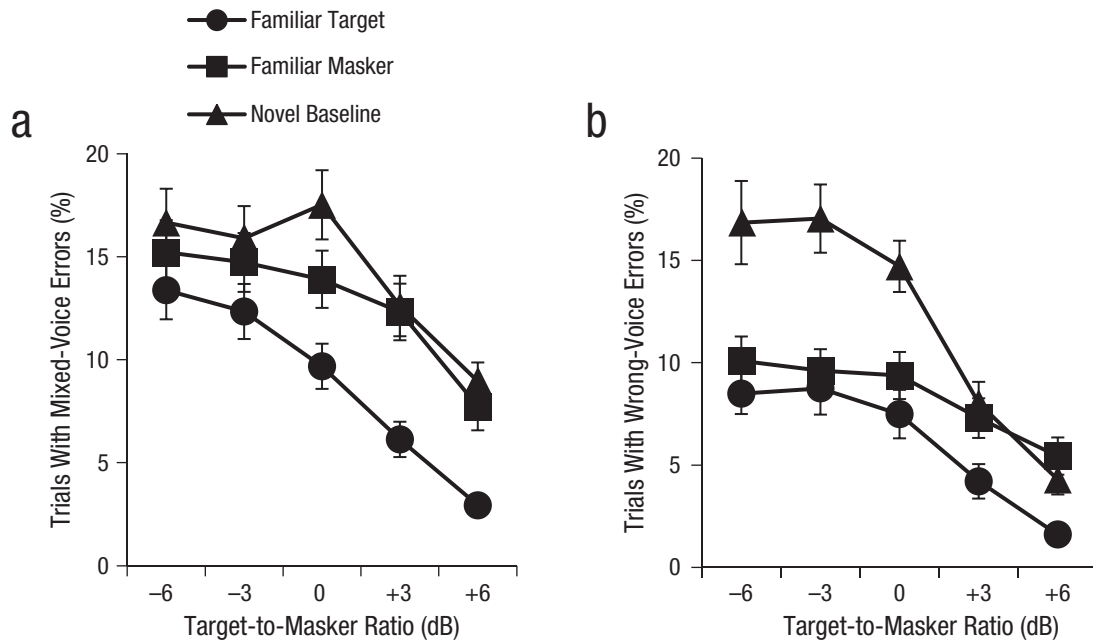
Fig. 5. Percentage of trials on which participants made (a) mixed-voice errors and (b) wrong-voice errors as a function of target-to-masker ratio and condition. Error bars show standard errors of the mean.

observed not only when the familiar voice was the target, but also when the familiar voice was to be ignored and a novel voice was to be tracked. Familiar-voice information may assist listeners in more than one way, because age correlated differently with performance in the familiar-target and familiar-masker conditions. Whereas the intelligibility of novel voices (masked with novel or familiar voices) declined with age, intelligibility of a familiar voice did not. This suggests that voice familiarity may help

older listeners compensate for sensory or cognitive decline.

"Sequential organization" refers to perceptual organization over time, as the sound unfolds. Sequential organization has been divided by Bregman (1990) into "primitive" processes dependent on acoustic cues (such as differences in frequency and timbre), and "schema-driven" processes that are dependent on knowledge of the to-be-reported sounds. The ability to exploit
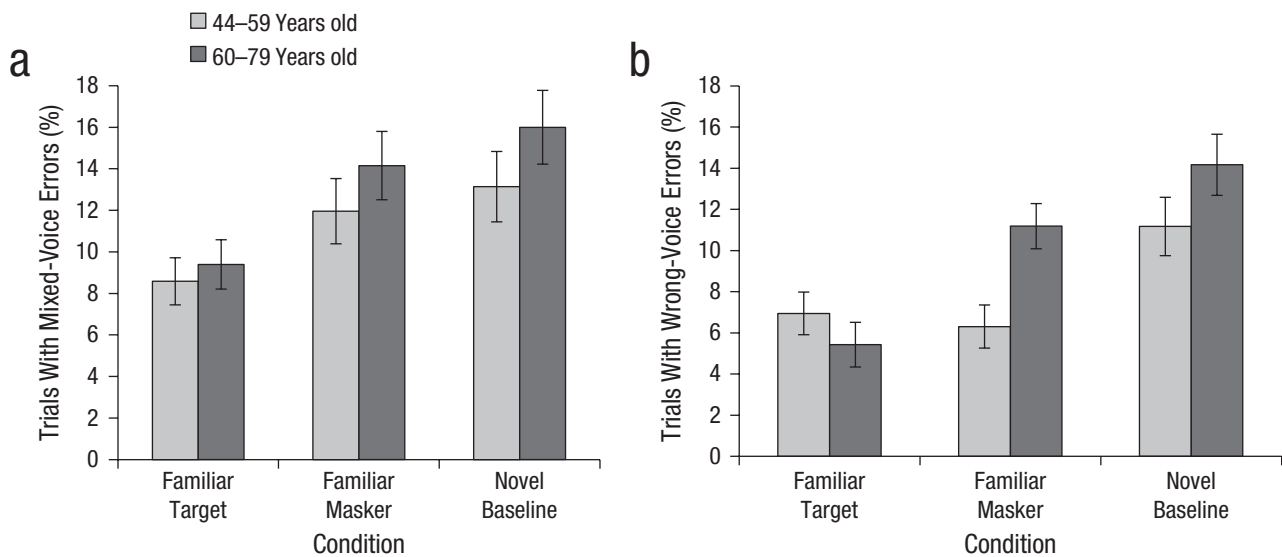

Fig. 6. Percentage of trials on which participants made (a) mixed-voice errors and (b) wrong-voice errors as a function of condition and age group, collapsed across target-to-masker ratio. Error bars show standard errors of the mean.

familiarity with a masker in this study contrasts markedly with the results of experiments in which listeners were required to identify a familiar melody when its notes alternated with those of another melody (Dowling, 1973; Hartmann & Johnson, 1991). Bregman (1990) observed that familiarity with a target melody, but not with a masker melody, helped listeners to identify a target melody, and he identified this asymmetry as a hallmark of schema-driven processes. In contrast, we found that familiarity with a masking voice facilitates perception. This difference may be related to the nature of the familiar material: The interleaved-melodies task involved familiarity with the melody itself, which is similar conceptually to a speech message, whereas we manipulated familiarity with a voice, which is more analogous to the instrument (i.e., source) carrying the melody. Our results also differ from those of a study by Newman and Evers (2007), who reported a weak effect of target familiarity, but no effect of masker familiarity, in a speech-shadowing task. Numerous differences between the two studies include a difference in degree of familiarity: having attended lectures by a university professor in their study versus many years of marriage in the present study. Our results demonstrate for the first time that familiarity can assist with the perceptual separation of target and masker voices, even when the familiar source is not the focus of attention and when listeners do not know in advance to which signal they will be expected to listen.

In the present study, sequential organization can be thought of as "tracking" the words spoken by one voice, from the call sign through the two coordinate keywords. Wrong-voice errors could result from a failure to segregate the two call signs, from a failure to group the target call sign with the corresponding color and number, or from a bias toward reporting a familiar voice. We can rule out this latter attentional explanation for the differences in errors between conditions because wrong-voice errors were reduced in the familiar-masker compared with the novel-baseline condition. In contrast, mixed-voice errors reflect a failure to link the two coordinate keywords spoken by the correct talker, unambiguously indicating a failure of sequential organization. Listeners made fewer errors of this type in the familiar-target than in the novel-baseline condition, which suggests that target familiarity facilitates segregation of two concurrent voices. Our results indicate that current accounts of sound segregation, which focus on stimulus-driven, automatic processes, need to be revised to accommodate the effects of knowledge. Listeners can benefit not only from familiarity with the target, but also from familiarity with the masker, and this result cannot be attributed to template matching, guessing, or a tendency to selectively attend to the familiar voice.

Although we cannot be certain which aspects of sound segregation are affected by familiarity, the size of the effects reported here are substantial. Performance in the familiar-target condition at a TMR of −6 dB fell between that in the novel-baseline condition at 0 and +3 dB, which indicates a 6 to 9 dB benefit. This knowledge-based factor facilitates performance nearly as much as do two acoustic factors that are well known to have a large influence on perceptual organization—whether the two voices are of the same or different sex (Brungart et al., 2001) and whether they come from the same or different locations (Hawley et al., 2004; Plomp & Mimpen, 1981). Our procedure probably underestimates the advantage to be gained from a familiar voice, because the CRM stimuli are stereotyped and impoverished in their content compared with the prosodic and contextual richness of everyday conversation, in which individual differences in voice quality (and in the resulting benefit) are likely to be substantially greater.

Our results confirm that the knowledge that older adults have gained about the voices of their friends and family members over the years is useful in multitalker environments. Unlike the intelligibility of novel voices, which declines with age (Figs. 3b and 3c), no age-related decline in the intelligibility of familiar-voice targets was observed (Fig. 3a), although listeners' performance was not apparently at ceiling. The familiarity manipulation did not affect the acoustics of the stimuli at all; in fact, our counterbalancing of novel and familiar voices ensured that, across the age group, the acoustic characteristics of all three conditions were matched. The apparent decline with age in the intelligibility of novel voices cannot be simply due to increased energetic masking resulting from age-related changes in frequency selectivity (van Rooij et al., 1989) and must therefore be related to nonauditory (i.e., cognitive) changes with age. Familiar-voice information appears to help listeners compensate for such age-related changes.

The role of knowledge-based factors in speech comprehension is likely to increase in importance, both as listeners age and as the acoustic environment becomes less favorable. Our results demonstrate two ways in which one knowledge-based factor—voice familiarity—may aid perceptual organization: Listeners can exploit the familiarity of both target and interfering voices to enhance intelligibility of a target signal. Younger listeners can exploit the familiarity of interfering and target voices, whereas, for older listeners, familiarity of the target voice may be of primary importance (as suggested by our error data). Who you choose to talk to at a cocktail party may have many consequences: Our results show that these include the intelligibility of a conversational partner in the presence of competing speech.

Author Contributions

Supplemental Material

Additional supporting information may be found at http://pss .sagepub.com/content/by/supplemental-data

References

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

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, *107*, 1065–1066.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustical Society of America*, *109*, 2276–2279.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *110*, 2527–2538.

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.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*, 414–427.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Handbook of perception and cognition: Hearing* (2nd ed., pp. 387–424). London, England: Academic Press.

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

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*, 85–99.

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

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America*, *115*, 833–843.

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

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

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

Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acustica*, *48*, 325–328.

van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, *121*, 519–526.

van Rooij, J. C., Plomp, R., & Orlebeke, J. F. (1989). Auditive and cognitive factors in speech perception by elderly listeners: I. Development of test battery. *Journal of the Acoustical Society of America*, *86*, 1294–1309.

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*, 88–99.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399–413.