Speech spoken by familiar people is more resistant to interference by linguistically similar speech

E. Holmes[1] & I. S. Johnsrude[1,2]

[1] The Brain and Mind Institute, University of Western Ontario, London, Ontario, N6A 3K7, Canada

[2] School of Communication Sciences and Disorders, University of Western Ontario, London, Ontario, London, N6G 1H1, Canada

Corresponding author: E. Holmes; E-mail: emma.holmes@ucl.ac.uk; Phone: +44 7597 967397; Mailing address: Wellcome Centre for Human Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, United Kingdom.

Abstract

Understanding speech in adverse conditions is affected by experience—a familiar voice is substantially more intelligible than an unfamiliar voice when competing speech is present, even if the content of the speech (the words) are controlled. This familiar-voice benefit is observed consistently, but its underpinnings are unclear: Do familiar voices simply attract more attention, are they inherently more intelligible because they have predictable acoustic characteristics, or are they more intelligible in a mixture because they are more resistant to interference from other sounds? We recruited pairs of native English participants who were friends or romantic couples. Participants reported words from closed-set English sentences (Oldenburg matrix test; HörTech, 2014) spoken by a familiar talker (the participant's partner) or an unfamiliar talker. We compared three masker conditions that are acoustically similar but differ in their demands: (1) English Oldenburg sentences; (2) Oldenburg sentences in a language incomprehensible to the listener (Russian or Spanish); and (3) unintelligible signal-correlated noise. We adaptively varied the target-to-masker ratio to obtain 50% speech reception thresholds. We observed a large (~5 dB) familiar-voice benefit when the target and masker were both English sentences. This benefit was attenuated (to ~2 dB) when the masker was in an incomprehensible language and disappeared when it was signal-correlated noise. These results suggest that familiar voices did not benefit intelligibility because they were more predictable or because they attracted greater attention; rather, familiarity with a target voice reduced interference from maskers that are linguistically similar to the target.

Keywords

Familiarity; speech; voice; attention; auditory; language

Introduction

In many everyday situations, we face the challenge of conversing when background noise is present. The ability to understand speech decreases as the level of background noise increases in intensity (e.g., Freyman, Balakrishnan, & Helfer, 2001; Rosen, Souza, Ekelund, & Majeed, 2013). Yet, we frequently encounter the voices of people we know and such knowledge substantially improves intelligibility when other sounds are present (Holmes, Domingo, & Johnsrude, 2018; Johnsrude et al., 2013; Kreitewolf, Mathias, & von Kriegstein, 2017; Magnuson, Yamada, & Nusbaum, 1995; Newman & Evers, 2007; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Souza, Gehani, Wright, & McCloy, 2013; Yonan & Sommers, 2000). In this paper, we explore the cognitive mechanisms underlying this robust benefit of talker knowledge to perception.

Naturally familiar voices, such as the voices of friends (Domingo, Holmes, & Johnsrude, 2019; Holmes et al., 2018) and spouses (Domingo et al., 2019; Johnsrude et al., 2013), are more intelligible than unfamiliar voices when competing speech is present. Johnsrude et al. (2013) presented participants with two sentences on each trial, which were spoken at the same time by two different talkers. Participants were asked to report words from the sentence that began with a particular cue word. Intelligibility was 10–15% better when the target sentence was spoken by the participant's spouse (married for more than 18 years) than by unfamiliar talkers of the same sex as the spouse. We call this the familiar-voice benefit. Two more recent studies (Domingo et al., 2019; Holmes et al., 2018) that used similar methods but tested friends (presumably less familiar than long-married spouses) showed a familiar-voice benefit of a similar magnitude. In all three studies (Domingo et al., 2019; Holmes et al., 2018; Johnsrude et al., 2013), familiar and unfamiliar target stimuli were acoustically identical over the group—voices were counterbalanced such that familiar voices were used as unfamiliar voices for other

participants. Therefore, the familiar-voice benefit cannot be due to systematic differences in the acoustics of familiar compared to unfamiliar voices.

When characterising the origins of intelligibility benefits, masking effects are typically separated into two categories. Masking that physically interrupts or occludes a target signal so that it is effectively not transduced at the periphery is called *energetic masking*. When both target and masker are audible, but are difficult to separate perceptually, *informational masking* occurs (Kidd, Mason, Richards, Gallun, & Durlach, 2007; Leek, Brown, & Dorman, 1991; Pollack, 1975). Given that previous studies carefully counterbalanced the acoustics of familiar and unfamiliar voices (e.g., Johnsrude et al., 2013; Domingo, Holmes & Johnsrude, 2019; Holmes, Domingo & Johnsrude, 2018), the familiar-voice benefit must be attributable to a release from informational masking. However, informational masking (and its release) is a catch-all term, conflating a variety of cognitive phenomena, and we do not yet understand precisely how a familiar voice improves intelligibility.

One possibility is that talker knowledge facilitates perceptual organization of two talkers—for example, due to better simultaneous or sequential grouping (Bregman, 1990; Darwin, 1997; Micheyl, Shamma, Elhilali, & Oxenham, 2010) of speech when one of the talkers in the mixture is familiar than when all talkers are unfamiliar. If the familiar-voice benefit arises from better perceptual segregation, a familiar voice should facilitate speech intelligibility when it is the masker, not only when it is the target. Although one previous study (Johnsrude et al., 2013) found evidence for a familiar masker benefit and concluded this was likely due to better stream segregation, they used a task with a low memory load—meaning that participants could potentially divide their attention between both talkers and retrospectively report words from the target sentence; other studies using different tasks have found no detectable benefit of familiarity with the masker voice (Domingo et al., 2019; Newman & Evers, 2007; see also Samson & Johnsrude, 2013, who found a benefit of a consistent masker talker in two- but not three-talker mixtures). What we can rule out, at least in the studies conducted so far, is

obligatory biasing of attention towards a familiar voice being the explanation for familiar-target benefit. If this were the case, performance would be poorer than baseline when the familiar voice was the masker (because listeners would be reporting it, not the target). This pattern of results has not (yet) been reported. Thus, it is unlikely that improved segregation is the sole, or even main, mechanism yielding familiar voice intelligibility benefit when the familiar voice is the target in a two-voice mixture.

Broadly speaking, two plausible explanations exist. The first is that individuals are more sensitive to target speech in a familiar voice, and this promotes its intelligibility regardless of the content of the masker. This could be because acoustic characteristics of familiar voices are more predictable: previous experience with a voice may allow listeners to better pick out spectrotemporal features that belong to it. For example, predictions about the dominant frequencies of a familiar voice might lead to narrower peripheral filters when that voice is attended (Green & McKeown, 2001; Schlauch & Hafter, 1991). Or, knowledge of a voice might permit better perceptual continuity (i.e. better grouping of speech across time). Possibly, familiar voices may simply attract more attention or are more motivating when they are the target than unfamiliar voices, increasing the likelihood that words will be reported more accurately when the target talker is familiar. All of these possibilities would lead to a similar familiar-voice benefit to speech intelligibility across different types of maskers that have different contents but which are similar in their acoustics.

The second possible explanation is that the familiar-voice benefit is due to reduced interference from a masker when a familiar, compared to unfamiliar, target voice is heard—for example, because familiar voices require fewer resources to process. According to this view, the magnitude of the benefit will differ among masker conditions that have different cognitive (e.g., attentional) demands, but are otherwise acoustically matched. Linguistic similarity of target and masker speech is well-known to affect attentional demand (e.g., see Treisman, 1964)—when listeners are instructed to report target speech, maskers are more interfering when they

are linguistically similar than dissimilar to the target. If familiar voices provide the greatest benefit to speech intelligibility when the masker is linguistically similar to the target, and provide less benefit when the masker is unintelligible, this would suggest that familiarity improves intelligibility in a mixture by reducing interference from attentionally demanding maskers.

Although previous studies investigating the familiar-voice benefit to intelligibility have used different maskers (Barker & Newman, 2004; Johnsrude et al., 2013; Kreitewolf et al., 2017; Newman & Evers, 2007; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Souza et al., 2013; Yonan & Sommers, 2000), these studies also used substantially different tasks and types of familiar voice (naturally familiar or trained). Therefore, the familiar-voice benefit is difficult to compare across these studies. Here, we systematically studied the familiar-voice benefit with different maskers that differed systematically in their linguistic content.

We recruited pairs of young adults who had known each other for 6 months or longer (friends or romantic couples). The familiar and unfamiliar voices were identical across the group (i.e. familiar voices were presented as unfamiliar to other participants). We presented sentences from the closed-set Oldenburg International Matrix test (HörTech, 2014), which all have the form "<Name> <verb> <number> <adjective> <noun>" (e.g., "Peter got five large desks"). Participants were cued to the sentence beginning with the name word "Peter" and reported the other four words from the target sentence, which could be spoken by a familiar or unfamiliar talker. We measured speech reception thresholds (SRTs) by varying the difference in level (i.e., the target-to-masker ratio; TMR) between a target sentence and a competing sound to estimate the 50% threshold for reporting sentences correctly. This ensured that any differences between the conditions could not be explained by differences in performance—because performance (thresholds) were measured at 50% intelligibility in all conditions. We compared SRTs for sentences spoken by familiar and unfamiliar talkers under three masking conditions: (1) a competing talker who spoke the same language (English) as the target (for which the familiar-voice benefit has been reported consistently), (2) a competing talker who spoke a different

language that was incomprehensible to the listener, and (3) unintelligible speech-spectrum noise convolved with the amplitude envelope of one of the sentences from the other two conditions. These maskers are matched in their spectrotemporal energy and amplitude modulation envelopes but impose different attentional demands (e.g., see Treisman, 1964).

If a familiar voice is perceived more accurately (e.g., because its acoustic characteristics are more predictable) or aids perceptual segregation, then a familiar voice should be more intelligible in all three conditions. Similarly, a familiar voice should be more intelligible in all three conditions if it attracts greater attention than unfamiliar voices. In contrast, a greater familiar-voice benefit in condition (1) than in (3) would suggest that familiar voices reduce interference when the masker is competing speech. Including a foreign-language masker in condition (2) enables us to examine whether similar linguistic (i.e. phonetic) information between the target and masker contributes to the familiar-voice benefit: performance in condition (2) would resemble that in condition (3) if the familiar-voice benefit only occurs in the presence of intelligible speech; it would resemble that in condition (1) if the familiar-voice benefit occurs for all speech-like sounds; and it would be intermediate if the familiar-voice benefit depends on the degree of linguistic similarity between the target and masker, which systematically affects the attentional demands of the masker (Dai, McQueen, Hagoort, & Kösem, 2017; Treisman, 1964). We chose two languages (Russian and Spanish) that participants did not understand, rather than one, to ensure that the benefit of a familiar talker in the presence of a foreign-language masker was not due to specific properties of any one language.

Materials and Methods

Participants

We recruited 9 pairs of participants (3 male, 15 female) who had known each other for 0.9–7.3 years (median = 2.9 years, interquartile range = 2.5) and who spoke regularly (> 3.5 hours per week). Pairs of participants were friends or romantic couples. Three were opposite-

sex pairs and 6 were same-sex (both female) pairs. One participant took part in the recordings, but never completed the experiment, leaving seventeen participants.

The familiar-voice benefit to speech intelligibility found in previous studies has a large effect size [f = 0.72 in Johnsrude et al. (2013) and f = 0.88 in Holmes et al. (2018)], and effects of this size should be detectable with power ~1.00 with 17 participants. A sample size of 17 is estimated to be sensitive to within-subjects effects of size f = 0.32 with 0.95 power (Faul et al., 2007), which means that this design will be sensitive to differences between Masker conditions that are smaller than the familiar-voice benefit observed in previous studies.

Participants were aged 20–28 years (median = 22.5 years, interquartile range = 5.4) and were native Canadian English speakers with no history of hearing difficulty. None of the participants had ever learnt to speak or understand Spanish or Russian. Participants had average pure-tone hearing levels better than 25 dB HL in each ear (at five octave frequencies between 0.5 and 4 kHz). The experiment was cleared by Western University's Health Sciences Research Ethics Board and informed consent was obtained from all participants.

Apparatus

The experiment was conducted in a single-walled sound-attenuating booth (Eckel Industries of Canada, Ltd.; Model CL-13 LP MR). Participants sat in a comfortable chair facing a 24-inch LCD visual display unit (either ViewSonic VG2433SMH or Dell G2410t).

Acoustic stimuli were recorded using a Sennheiser e845-S microphone connected to a Steinberg UR22 sound card (Steinberg Media Technologies). During the listening tasks, acoustic stimuli were presented through the Steinberg UR22 sound card and were delivered binaurally through Grado Labs SR225 headphones.

Design

Participants completed two tasks: a speech intelligibility task and an explicit recognition task. Nine completed the speech intelligibility task first and 8 completed the explicit recognition task first.

In the speech intelligibility task, listeners heard an English target sentence on every trial, which was of the form "<Name> <verb> <number> <adjective> <noun>" (HörTech, 2014; Zokoll et al., 2013; see Table 1). An example is "Peter got three large desks". The target sentence always began with the name word "Peter". Participants identified the four remaining words of the target sentence by clicking buttons on a screen, which were arranged in a 4 x 8 matrix (Figure 1; chance rate: 0.02%). On each trial, the target sentence was spoken either by the participant's partner ("Familiar" condition) or by one of two unfamiliar talkers ("Unfamiliar" condition) who were the familiar talkers of other participants in the experiment.

Listeners also heard a masking stimulus on every trial, which was presented simultaneously with the target sentence. The masking stimulus was either a different English sentence spoken by an unfamiliar talker ("Native" condition), a different-language sentence (Spanish or Russian) spoken by an unfamiliar talker ("Foreign" condition), or unintelligible speech spectrum noise with the amplitude envelope of one of the sentences (i.e., signal correlated noise; "SCN" condition).

Each participant's familiar voice was presented as unfamiliar to two other participants in the experiment, which ensured that Familiar and Unfamiliar conditions were as acoustically similar as possible across the group. We asked participants to verify that the unfamiliar talkers they heard were unknown to them. We required four additional talkers who were not participants in the experiment to record the Spanish and Russian sentences. To ensure that acoustic differences between the voices of these talkers and the voices of our participants could not explain different results between the Native and Foreign masker conditions, the talkers who recorded the Spanish and Russian sentences (who were all bilingual) also recorded English

sentences. We presented these English sentences as maskers in a sub-set of trials in the Native condition to ensure that differences between talkers could not explain differences between masker conditions.

The aim of the explicit recognition task was to check whether participants could discriminate their partner's voice from the unfamiliar voices they heard in the experiment. On each trial, listeners heard one English sentence. The sentence was spoken by the participant's partner or by one of the four unfamiliar talkers they had heard (or would hear) in the speech intelligibility task. Participants reported whether each sentence was spoken by their partner or not, completing 95 trials (19 spoken by each of the five talkers).

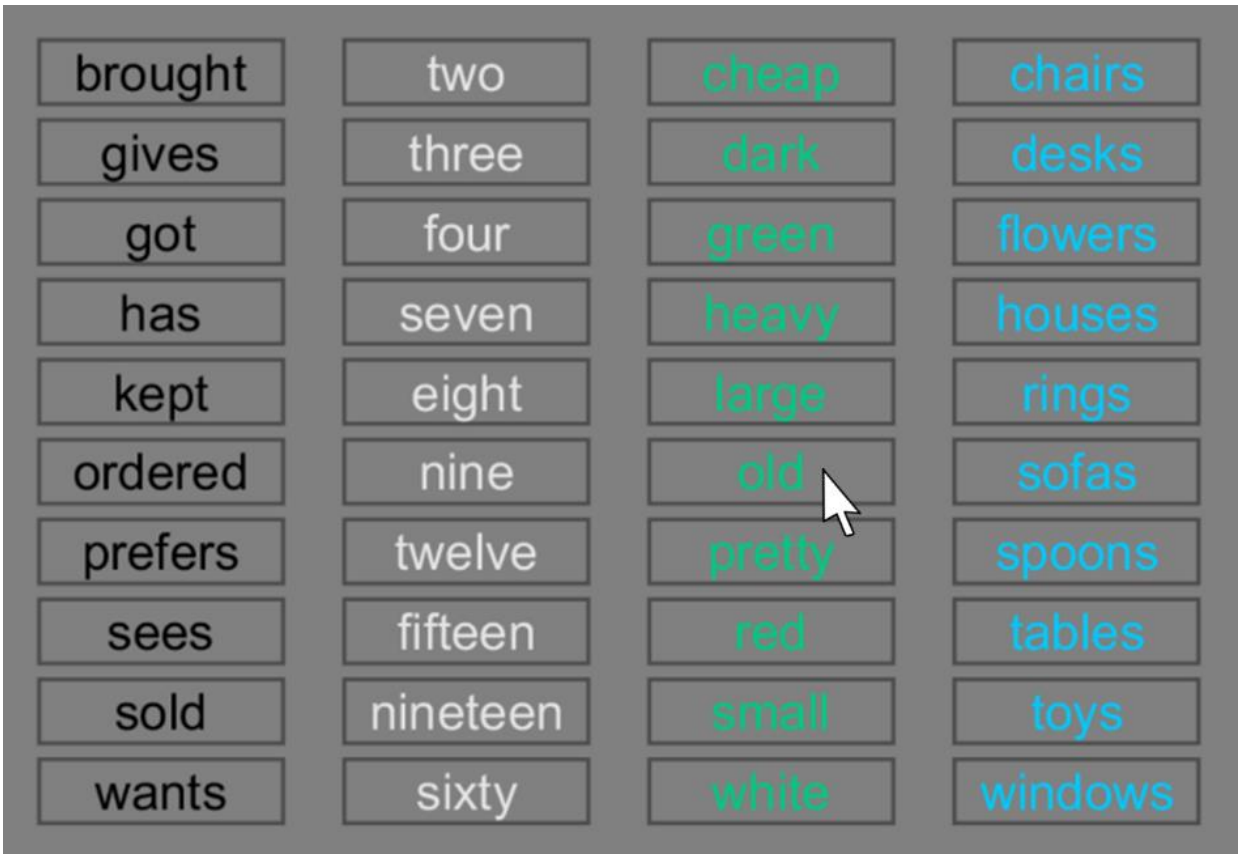

Figure 1. Response screen in the speech intelligibility task. On each trial, participants clicked one word from each column of buttons.

Stimuli

Each participant recorded 320 sentences from the English version of the Oldenburg International Matrix corpus (HörTech, 2014), described in the previous section. Sentences were recorded before each participant completed the listening part of the experiment, which always took place in a separate session on a different day.

< Insert Table 1 >

Four bilingual talkers, who were not participants in the experiment, recorded 160 English sentences (the subset of the English sentences that were presented as maskers) and 160 sentences from a different-language version of the Oldenburg International Matrix corpus. Two English-Russian bilingual talkers (1 male, 1 female) recorded sentences from both the English and the Russian (Warzybok et al., 2015; see Table 2) versions of the Oldenburg International Matrix corpus. The other two English-Spanish bilingual talkers (1 male, 1 female) recorded sentences from both the English and the Spanish (Hochmuth et al., 2012; see Table 3) versions of the Oldenburg International Matrix corpus. All four talkers were unfamiliar to participants; had lived in Canada for 12 years or longer; had completed most of their schooling in English; and used Spanish or Russian at home. Thus, they commanded good spoken language in Spanish and Russian, and had typical Canadian accents when speaking English.

< Insert Tables 2 & 3 >

To ensure that all sentences were spoken at similar rates, so that the five words from two different sentences would overlap when used in the speech intelligibility task, we played videos (Holmes, 2018) indicating the desired pace for each sentence while participants

completed the recordings. Participants were instructed to speak each word at the same time that a vertical bar passed the beginning of the written word, but were otherwise asked to speak the sentences as naturally as possible. The recorded sentences had an average duration of 2.8 seconds (standard deviation [*s*] = 0.3). The levels of the digital recordings of the sentences were normalised for root mean square (RMS) power.

Signal-correlated noise (SCN) was generated for each talker by calculating the long-term average spectrum of all sentences (for the bilingual talkers, we used only the Spanish or Russian sentences). A noise with this spectrum was then convolved with the amplitude envelope of each sentence. In other words, the periodic content of the noise was equivalent to the sentences in the other conditions. This method produced 320 SCN stimuli for each participant that corresponded to their English sentences, and 160 SCN stimuli for each of the 4 bilingual talkers, for the Russian and Spanish sentences.

Procedure

On each trial of the speech intelligibility test, a target sentence was presented at the same time as a masking stimulus. It was always in a different voice (or, for SCN, generated from a different voice) than the target. If it was English, we ensured the masker words were different from the target words. The target-to-masker ratio (TMR) started at 0 dB and was varied adaptively in a one-up one-down procedure (Wetherill & Levitt, 1965). After applying the TMR manipulation, the root-mean-square (rms) level of the combined sentence and masker was normalized. If the participant reported any of the words incorrectly, this was categorized as an incorrect trial. The adaptive procedure converged on the 50% threshold. The step size was 2 dB for the first three reversals, then 0.5 dB thereafter. The procedure stopped after 12 reversals and the speech reception threshold (SRT) for each run was calculated as the median of the last 6 reversal values. Participants received a break every 48 trials.

To measure all combinations of Familiarity (Familiar or Unfamiliar) and Masker (Native, Foreign, or SCN) factors for the different talkers, we adapted the TMR in 26 separate but

interleaved runs (ordered randomly), which are listed in Table 4 and labelled A–Z. Every run contained 160 possible target sentences and 160 possible masker sentences. The target sentence was always an English sentence, which could be spoken by the participant's familiar partner or by the partners of two other participants in the experiment, who were unfamiliar to the participant but were the same sex as their partner. In the Native masker condition (runs A–J), the masker was a different English sentence, in which all five words were different from the target words. Four different masker voices were used in separate runs (except that a voice was never used as a masker if it was also the target). These masker voices included the same two unfamiliar talkers that were presented as targets and the Spanish and Russian bilingual talkers who were the same sex as the listener's partner. In trials of the Foreign masker condition (runs K–P), the masker was a Spanish or Russian sentence spoken by the same two bilingual talkers (half of the runs with each masker voice). In the SCN masker condition (runs Q–Z), noise was used as a masker. Noise maskers were generated from sentences spoken by the same four unfamiliar voices that were used as maskers in the other conditions. Noise maskers were generated from English sentences from the unfamiliar talkers that provided maskers in the Native condition, and from foreign-language sentences spoken by bilingual talkers that were maskers in the Foreign condition. Although we did not expect the results to differ systematically when these different voices were used as maskers, or between conditions in which SCN was generated from different language sentences, this design minimised the possibility that differences between Native, Foreign, and SCN conditions could be explained by differences in acoustics. The median number of trials for each participant, across all runs, was 756.0 (interquartile range = 94.3).

< Insert Table 4 >

Analyses

For the explicit recognition task, we calculated sensitivity (d'; Green & Swets, 1966) using loglinear correction (Hautus, 1995). Chance d' was 0.3 and the maximum attainable d' was 4.4.

For the speech intelligibility task, we calculated the mean SRTs across runs for each combination of Familiarity (2 levels) and Masker (3 levels) factors (see Table 4). To evaluate the effects of these factors and their interactions on SRTs, we used a two-way repeated-measures ANOVA and posthoc comparisons.

To examine the effect of familiarity on behaviour in more detail, we analysed data from error trials in the Native masker condition. When listeners did not report the entire sentence correctly (classified as incorrect for the adaptive procedure), we calculated the percentage of words that were correct ('Target' words) and compared these percentages between the Familiar and Unfamiliar conditions using a paired-samples *t*-test. Words that were not from the target sentence could either be from the masker ('Masker' words), or not present in either sentence ('Random' words). We compared the percentages of these Word Types (Masker and Random) across Familiarity (2 levels) in a repeated-measures MANOVA. If voice familiarity does not influence the type of errors participants make, we would expect to find no effect of familiarity on Masker or Random Word Types.

Results

Results from the yes-no Explicit Recognition task confirmed that participants could recognise their partner's voice with near-perfect sensitivity when presented in quiet (median d′ = 4.4, interquartile range = 0.0, range = 3.9–4.4).

Greatest benefit from familiar voice for same-language masker

We analysed the Speech Intelligibility data using a 2 x 3 within-subjects ANOVA with the factors Familiarity (Familiar and Unfamiliar) and Masker (Native, Foreign, and SCN). We first

confirmed that the speech intelligibility data met the assumptions of normality, as assessed by combining evidence from the Shapiro-Wilk test with observations of histograms, box-plots, and Q-Q plots. The only exception was the SCN conditions, for which the assumption of normality was violated because one participant was an outlier (their threshold was more than 1.5 times the interquartile range from the upper quartile of the group). We conducted the analyses in the SCN conditions with and without this participant and found the same pattern of results, so we report results with this participant included here.

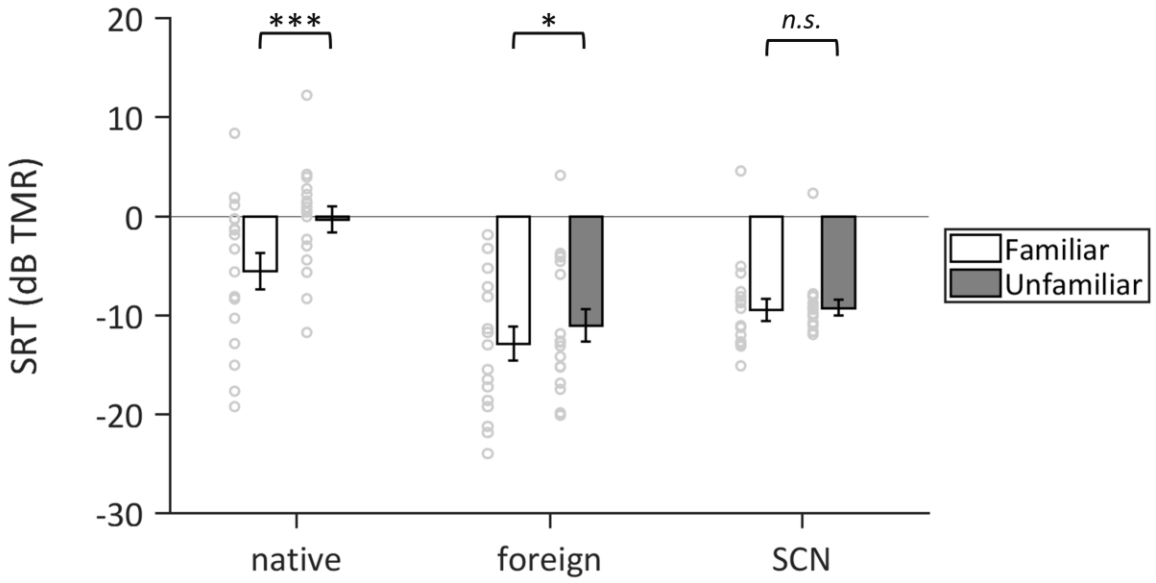

Figure 2. Speech reception thresholds (SRTs), expressed as target-to-masker ratio (TMR) in decibels (dB), across the two familiarity conditions (Familiar and Unfamiliar) and three masker conditions (Native, Foreign, and Signal-Correlated Noise [SCN]). Error bars represent ± 1 standard error of the mean. Dots indicate performance for individual participants. Brackets display the significance level of pairwise comparisons (* $p < .050$; ** $p < .010$; *** $p < .001$; *n.s.* not significant).

The main effect of Familiarity was significant, with better performance in Familiar than Unfamiliar conditions [$F(1, 16) = 13.37$, $p = 0.002$, $\omega_p^2 = 0.41$] (see Figure 2). We found a significant main effect of Masker Type [$F(2, 32) = 38.85$, $p < 0.001$, $\omega_p^2 = 0.68$]. Post-hoc tests with Bonferroni correction for multiple comparisons showed that SRTs were significantly better (i.e., lower) in the Foreign [$t(16) = 9.49$, $p < 0.001$, $d_z = 2.30$] and SCN [$t(16) = 6.23$, $p < 0.001$, $d_z = 1.51$] conditions than in the Native condition. There was a non-significant trend towards better SRTs in the Foreign than SCN condition [$t(16) = 2.24$, $p = 0.12$, $d_z = 0.54$].

There was a significant interaction between Familiarity and Masker [$F(2, 32) = 10.14$, $p < 0.001$, $\omega_p^2 = 0.34$]. SRTs were significantly better for familiar-voice than unfamiliar-voice targets in the Native masker condition (mean difference [$\bar{x}$] = 5.2 dB, standard deviation of difference [s] = 4.9) [$t(16) = 4.43$, $p < 0.001$, $d_z = 1.07$] and in the Foreign masker condition (mean [$\bar{x}$] = 1.9 dB, $s = 3.4$) [$t(16) = 2.24$, $p = 0.040$, $d_z = 0.54$], but not in the SCN condition ($\bar{x}$ = 0.2 dB, s = 3.0) [$t(16) = 0.31$, $p = 0.76$, $d_z = 0.08$]. If we correct for multiple (3) comparisons using Bonferroni correction, the difference in the Native condition remains significant ($p = 0.001$), the difference becomes non-significant in the Foreign masker condition ($p = 0.12$), and it remains non-significant in the SCN condition ($p \approx 1.00$). Thus, we found a familiar-voice benefit in the

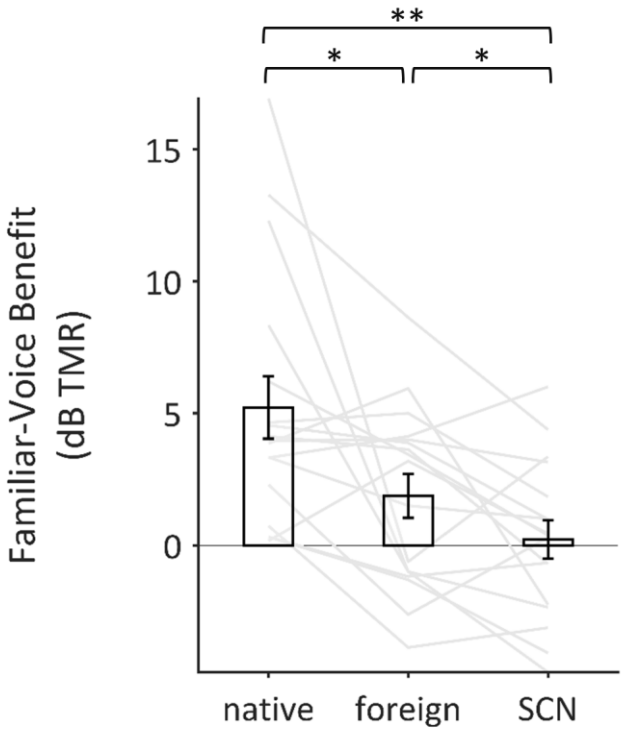

Figure 3. Familiar-voice benefit in the three masker conditions (Native, Foreign, and Signal-Correlated Noise [SCN]), expressed as the difference in speech reception thresholds (in dB TMR) between the two familiarity conditions (Familiar and Unfamiliar). Error bars represent ± 1 standard error of the mean. Grey lines display the familiar-voice benefit for each participant (N = 17). Brackets display the significance level of pairwise comparisons (* $p < .050$; ** $p < .010$; *** $p < .001$; *n.s.* not significant).

Native condition, a small benefit around the significance threshold in the Foreign condition, and no benefit in the SCN condition; the difference in the familiar-voice benefit among Masker conditions is supported by a significant interaction.

Given our hypothesis was about the magnitude of the familiar-voice benefit among the masker conditions, and the interaction was significant, we calculated the magnitude of the Familiar-Voice Benefit (Familiar – Unfamiliar condition), and compared it between Masker

conditions using paired-samples *t*-tests (planned comparison). Figure 3 plots the magnitude of the benefit for each Masker condition. We found significant differences in the magnitude of the benefit between all three pairs of maskers: it was larger in the Native condition than in the Foreign [$t(16) = 2.52$, $p = 0.023$, $d_z = 0.61$] and SCN [$t(16) = 4.04$, $p = 0.001$, $d_z = 0.98$] conditions, and larger in the Foreign condition than in the SCN condition [$t(16) = 2.23$, $p = 0.040$, $d_z = 0.54$].

Familiar voice reduces interference from masker talker

To understand more about the familiar-talker benefit in the Native condition, we analysed the words that participants reported on incorrect trials (i.e., trials in which they reported one or more words incorrectly). The percentage of words that belonged to the Target, Masker, and Random categories are illustrated in Figure 4.

First, we compared the percentage of Target words between the Familiar and Unfamiliar conditions. Participants reported a significantly greater percentage of Target words when the target sentence was spoken by the familiar talker than when it was spoken by one of the unfamiliar talkers [$t(16) = 4.93$, $p < 0.001$, $d_z = 1.20$]. In other words, when target sentences were spoken by a familiar talker, not only did listeners report more sentences correctly (as documented above), but they also reported more words correctly from sentences in which they made at least one error.

Next, we compared the percentages of Masker and Random (error) words across Familiarity conditions. A MANOVA revealed a significant effect of Familiarity [$F(1,16) = 17.21$, $p < 0.001$, $\omega_p^2 = 0.47$]. Follow-up univariate analyses with Bonferroni correction showed that participants made significantly fewer Masker errors on Familiar than Unfamiliar trials [$F(1,16) = 34.37$, $p < 0.001$, $\omega_p^2 = 0.65$], but there was no difference in the percentages of Random errors [$F(1,16) = 2.61$, $p = 0.25$, $\omega_p^2 = 0.08$]. A paired-samples *t*-test confirmed that the difference in the percentages of Masker words between Familiar and Unfamiliar trials was significantly greater than the difference in Random words [$t(16) = 4.55$, $p < 0.001$, $d_z = 1.10$]. These results

indicate that improved intelligibility of a familiar-voice is due (in part) to a lower probability of reporting words from a native-language masker, consistent with results reported by Domingo et al. (2019).

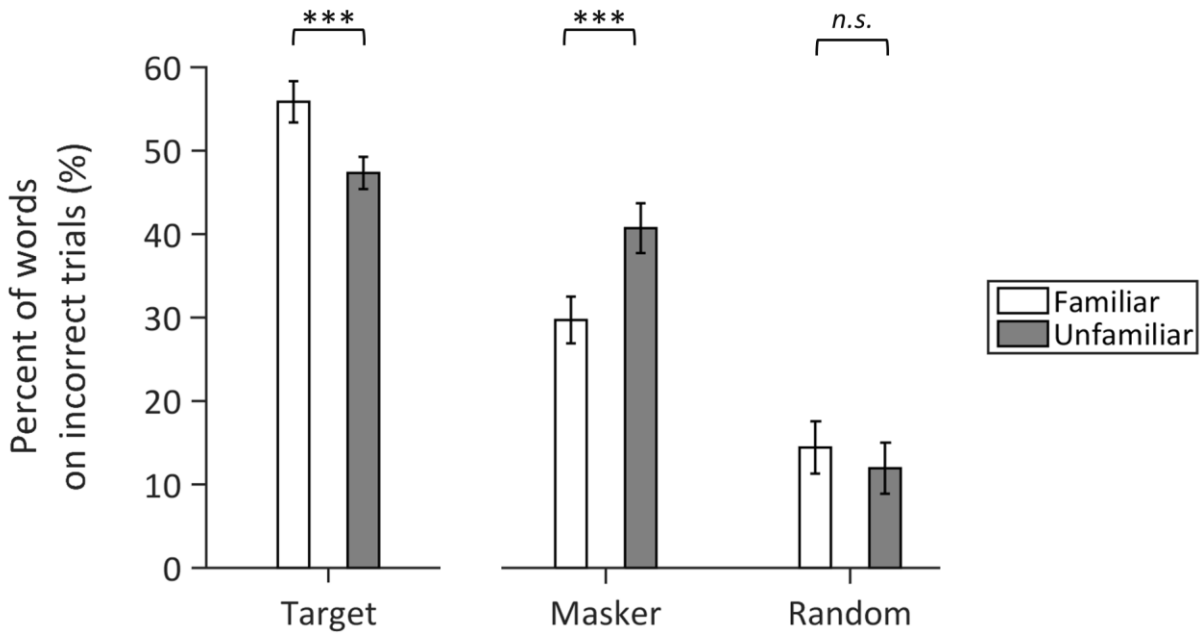

Figure 4. Analysis of words reported on incorrect trials (in which at least one word of four was reported incorrectly) in the Native Masker condition. “Target” words were correct (i.e. spoken by the target talker). “Masker” words were spoken by the masking talker (who was always unfamiliar). “Random” words were not present in either the target or masking sentence. The percentage of each type was calculated by dividing the number of words in each category by the total number of words on incorrect trials for each participant. Error bars represent one standard error of the mean. Brackets display the significance level of pairwise comparisons (* $p < .050$; ** $p < .010$; *** $p < .001$; *n.s.* not significant).

No difference in the familiar-talker benefit between Russian and Spanish maskers

We conducted a 3-way within-subjects ANOVA to examine: (1) whether we found the same effects on speech intelligibility when the voices that were used as maskers in the English and Foreign conditions were identical, and (2) whether the familiar-voice benefit differed between the two foreign languages we used (i.e., Spanish and Russian). The ANOVA included the factors Familiarity (Familiar and Unfamiliar), Masker (Native and Foreign), and Talkers (Spanish bilinguals and Russian bilinguals) and used the data from Runs C, D, F, G, and I–P (see Table 4). The data are displayed in Figure 5.

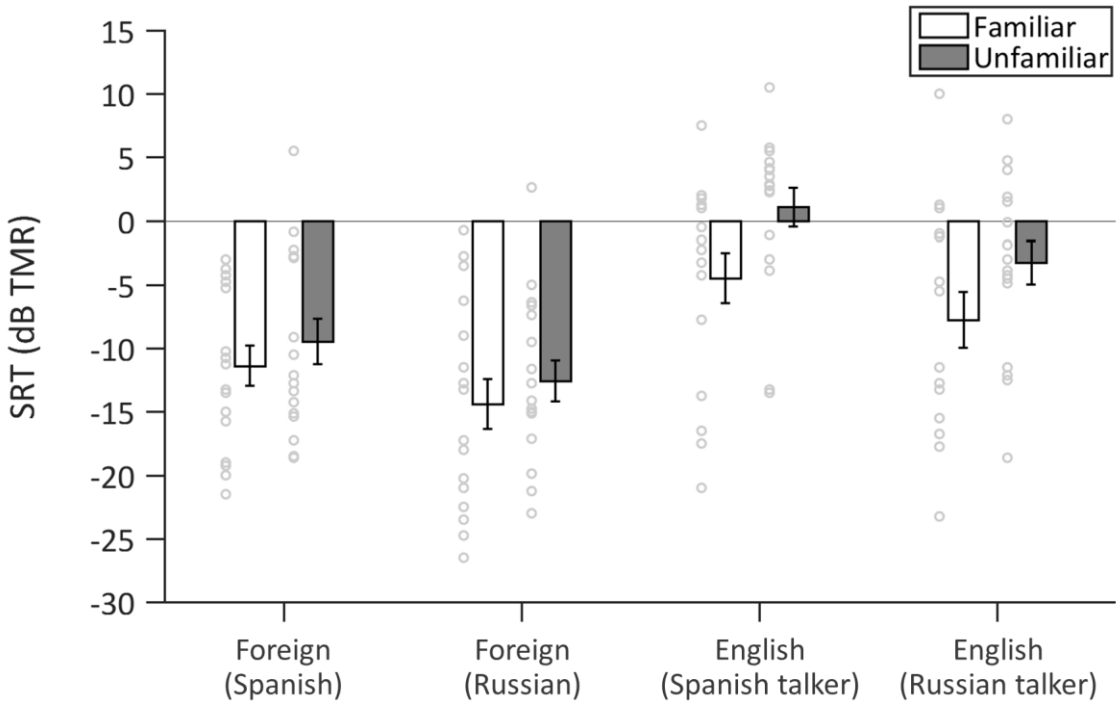

Figure 5. Speech reception thresholds (SRTs), expressed as target-to-masker ratio (TMR) in decibels (dB) for the two bilingual talkers whose voices were used as maskers in both the Foreign and English masker conditions. Error bars represent one standard error of the mean. Dots indicate performance of individual participants.

Consistent with the results reported above (in the section entitled “Greatest benefit from familiar voice for same-language masker”), SRTs were better when the target was Familiar than Unfamiliar [$F(1, 16) = 22.33$, $p < 0.001$, $\omega_p^2 = 0.18$]. We found a strong trend towards a significant interaction between Familiarity and Masker [$F(1, 16) = 4.30$, $p = 0.055$, $\omega_p^2 = 0.15$], demonstrating a trend towards a larger familiar-voice benefit to intelligibility in the English masker condition. It is worth noting that the trials that contributed to the English condition of this analysis included only 60% of the Runs that were used in the analyses reported in the previous section (entitled “Greatest benefit from familiar voice for same-language masker”), so we would expect these data to be noisier.

To investigate differences between the two languages, we looked at main effects and interactions with the Talker factor. The Russian talkers yielded lower SRTs than Spanish talkers [$F(1, 16) = 16.87$, $p < 0.001$, $\omega_p^2 = 0.47$]. Importantly, however, the two-way interaction between Talker and Familiarity was not significant [$F(1, 16) = 0.24$, $p = 0.63$, $\omega_p^2 = -0.04$]: the improvement in SRT gained from a familiar voice was similar regardless of the language (Spanish or Russian) and voice of the masker. The two-way interaction between Talker and Masker was not significant either [$F(1, 16) = 0.17$, $p = 0.68$, $\omega_p^2 = -0.05$], and neither was the three-way interaction between Talker, Masker, and Familiarity [$F(1, 16) = 0.12$, $p = 0.73$, $\omega_p^2 = -0.05$]. The presence of a main effect of Talker, but absence of an interaction, is consistent with the idea that participants performed better when the Russian bilinguals spoke the masker sentence regardless of whether they spoke an English or Russian sentence; therefore, the Russian talkers, rather than the Russian language, were likely responsible for the differences in SRTs between the Russian and Spanish talkers.

Overall, these results indicate that the familiar-voice benefit did not differ significantly between Spanish and Russian maskers, or between talkers.

Discussion

The magnitude of the familiar-voice benefit—measured as the difference in intelligibility for a familiar compared to unfamiliar voice in the presence of a masker—was largest when the masker was a same-language sentence (~5 dB TMR), smaller when it was a different-language sentence (~2 dB TMR), and undetectable when it was unintelligible SCN. Pairwise comparisons indicated that the magnitude of the familiar-voice benefit differed significantly among all three masker conditions: thus, the familiar-voice benefit does not appear to be all-or-none, but is instead graded. Our results are consistent with the idea that the magnitude of the familiar-voice benefit depends on the linguistic content of the masker, since the acoustic properties were as similar as possible: the long-term average spectrum and amplitude envelope were matched across all three masker conditions. The pattern of findings is compatible with the explanation that familiar voices reduce interference from masking sounds in attentionally demanding conditions when the masker is linguistically similar to the target.

The finding of better intelligibility for familiar than unfamiliar talkers in the presence of a competing same-language sentence (Native condition) is consistent with previous work (Domingo et al., 2019; Holmes et al., 2018; Johnsrude et al., 2013; Newman & Evers, 2007). However, the magnitude of the intelligibility benefit has never been directly compared under different masking conditions. We found the familiar-voice benefit was significantly smaller when the masker was in an incomprehensible language and disappeared entirely when the masker was unintelligible noise.

The finding that the familiar-voice benefit differed between masker conditions indicates that the benefit is unlikely to arise from processes that operate simply upon the familiar voice; for example, due to more precise predictions of the acoustic characteristics of a familiar than unfamiliar voice, or the ability to match spectrotemporal features to stored memories of a familiar voice. In addition, if familiar voices induce greater attention, motivation, or arousal than unfamiliar voices, then we would expect to find a familiar-voice benefit in all three conditions.

Instead, our pattern of results demonstrate that familiar-voice effects interact with the masker, and is not present when the masker is noise that contains no linguistic content.

Overall, TMRs corresponding to the 50% SRT were higher (indicating poorer performance) when the masker was the same language as the target compared to when it was a different language or unintelligible noise. However, the SRT in the Native condition was significantly lower (better), and more like that for the SCN condition, when the target sentence was spoken by a familiar than unfamiliar talker (see Figure 2). This finding implies that linguistic similarity between the target and masker sentences interfered with word report in the Native condition, but presenting target speech in a familiar voice reduced this interference.

The pattern of familiar-voice benefits across the three types of masker is consistent with the idea that a familiar talker reduces informational masking by reducing interference from linguistically similar sounds. When the masker was in an incomprehensible language, the familiar-voice benefit was smaller than when it was comprehensible, but larger than when it was unintelligible noise—consistent with the pattern expected based on the attentional demands of these maskers (e.g., Treisman, 1964). A same-language masker contains the most linguistic information, and is most confusable with the target, since masking words are potential target words. It seems reasonable to assume that this condition would be the most cognitively demanding. A different-language masker contains less relevant linguistic information, although it may contain phonemes similar to English. From the perspective of computational linguistic models (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris & McQueen, 2008), these phonemes could ‘activate’ lexical representations of English words, similar to phonetic priming of words by nonwords that contain overlapping phonemes (e.g., Slowiaczek, Nusbaum, & Pisoni, 1987). These different-language words can be easily discounted as belonging to the masker after the word structure diverges from possible English words, meaning that they cannot be target words. Therefore, they should interfere with target speech less than same-language maskers. In contrast, SCN contains no linguistic information, so is unlikely to activate competing

English words and should not cognitively interfere with the target sentence. If familiar voices only helped participants to better select which voice to attend to when both target and masker *words* (not just sounds) were possible targets (i.e. English words), then we should have found a familiar-voice benefit in the same-language condition but in neither of the other two conditions—which is different to the pattern of results we observed. The fact that we see some benefit for unintelligible (Spanish or Russian) speech suggests that interference on a phonological (as well as lexical) level may be reduced when a voice is familiar.

The pattern of errors also supports the idea that familiar voices help listeners to resist interference from linguistically similar competing speech. On Native masker trials when listeners did not report a target sentence correctly, they tended to report more words from the target sentence and fewer words from the masker sentence when the target was familiar, consistent with the idea that they were less distracted by the masker. In contrast, there was no difference between familiar and unfamiliar trials in the proportion of Random errors—i.e., incorrect words that were not spoken by either talker. Several previous studies show the ability to inhibit distracting information is important for tasks in which multiple talkers speak simultaneously (Melara, Rao, & Tong, 2002; Perrone-Bertolotti, Tassin, & Meunier, 2017; Treisman, 1964), and the current findings are compatible with the idea that a familiar target voice helps to reduce such interference. If the familiar-voice benefit was biggest in the Native condition because the acoustic characteristics of the masker differed from the other maskers, then we would expect the proportions of Random and Masker errors to be similar on familiar and unfamiliar trials—which is not what we found.

It is unlikely that the familiar-voice benefit that we observed is due to improved perceptual segregation, since this would also manifest across all three conditions and not only in the linguistically similar condition. Furthermore, if familiar voices enhance segregation, a benefit should be evident when the familiar voice is the masker as well as the target: Domingo et al. (2019) observed a familiarity benefit only when the target—not the masker—was familiar.

We therefore think it is unlikely that our results could be explained by better perceptual segregation—even if we were to assume that the benefit to intelligibility from perceptual segregation varies with acoustic and linguistic properties of a masker.

Simple familiarity of masker content cannot be the fully story either, because several training studies have shown that the intelligibility benefit for familiar voices transfers to masking settings other than the one in which the voice was learned (e.g., Holmes et al., 2019; Nygaard & Pisoni, 1998; Nygaard et al., 1994); for example, Holmes, To, & Johnsrude (in prep) show that new voices which are trained to become familiar produce a familiarity benefit when presented with a single competing talker, irrespective of whether they were learned in quiet or babble noise. These studies demonstrate that a familiar-voice benefit is observed even with the listener has no experience hearing that familiar voice in the presence of the experimental masker.

To investigate the reduction from interference in more detail, future work could investigate other conditions under which cognitive demand might affect the familiarity benefit, such as the cognitive demand from a secondary auditory task. Alternatively, providing visual linguistic interference—in the form of English words, foreign words, or non-linguistic visual stimuli—would help to explore the question of whether the reduction of interference from a familiar voice is specific to auditory interference or whether it arises from domain-general linguistic processes.

SRTs were lower (better) overall in the Foreign than Native condition, consistent with previous observations (Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; Calandruccio, Dhar, & Bradlow, 2010; Freyman, Helfer, McCall, & Clifton, 1999; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). When masking speech is incomprehensible to the listener, there is reduced competition of linguistic information between the target and masker (Dai et al., 2017; Van Engen & Bradlow, 2007), which enables participants to achieve lower (better) thresholds. In an experiment in which participants were trained to comprehend noise-vocoded speech, Dai et al. (2017) observed that participants reported more words from a target sentence

correctly before they had been trained to understand a vocoded speech masker (i.e. when it was unintelligible) than after they had been trained to understand the same masker as intelligible. This finding corroborates the idea that unintelligible speech produces less informational masking than intelligible speech, using acoustically identical stimuli.

If the familiar-voice benefit related only to the level of the masker or to the audibility of the target (for example, because familiar-voice information is more difficult to pick out at lower SNRs), then we would expect to observe the greatest benefit in the English masker condition, a smaller benefit in the SCN condition, and the smallest benefit in the Foreign condition, based on the average thresholds from those conditions. Instead, we found a larger benefit in the Foreign condition than the SCN condition, for which we found no familiar-voice benefit. We can also rule out an explanation based on difficulty because all of the conditions were adapted to the participant's 50% threshold, so accuracy was equated across the conditions.

This finding of no familiar-voice benefit when the masker was SCN differs from previous observations of a familiar-voice benefit in white (Nygaard & Pisoni, 1998; Nygaard et al., 1994) and signal-correlated (Kreitewolf et al., 2017; Souza et al., 2013) noise. However, most of these studies (Nygaard & Pisoni, 1998; Nygaard et al., 1994; Souza et al., 2013) used open-set tests, in which participants freely reported the words they heard, rather than having a set of options to select from. In open-set tests using everyday sentences, a greater tendency to guess words spoken by familiar than unfamiliar talkers (i.e. bias)—coupled with the semantic constraints inherent in meaningful, well-formed sentences—could artificially inflate word report for materials spoken by familiar talkers: first, because words are predictable and guesses are likely to be correct and, second, because responses are not constrained and participants can guess as many words as they choose; guessing more words is likely to result in more correct responses. Whereas, in the current closed-set task, participants always reported exactly 4 words on every trial, and guesses were unlikely to be correct. Kreitewolf et al. (2017) did use a closed-set test and observed significantly better SRTs for familiar than unfamiliar voices in SCN. However, the

benefit they found was very small (0.52 dB)—much smaller than the familiar-voice benefit we observed here—and could be due to familiarizing participants with both voices *and* maskers during the training portion of their experiment. During training, their participants had to identify words spoken by a talker in the presence of SCN, and the same SCN background was used in their speech intelligibility task. Possibly, the small familiar-voice benefit they observed might reflect learning specific to the trained masker and may not generalize to new listening environments. This differs from naturally familiar voices, like those used here, which are encountered in a wide variety of acoustic settings.

The relationship we observed between the familiar-voice benefit and the attentional demands of the masker may arise because phonological/lexical information is retrieved differently for familiar and unfamiliar talkers. The episodic account of speech recognition (Goldinger, 1996, 1998) proposes that each instance of a word is stored as an episodic memory and new words are recognized by comparing the acoustic signal against these stored memories. Under this account, words spoken by an unfamiliar talker will activate episodic traces less strongly than words spoken by a familiar talker: the acoustic input will match episodic traces less well. Thus, words spoken by an unfamiliar talker must be recognized through a normalization and matching process, which is slow. Exposure to someone's voice increases the number of their words that are stored in memory. When listening to a word spoken by a highly familiar talker, the acoustic signal will strongly activate similar episodes that are stored in memory, enabling fast word recognition. Under this account, recognition of words spoken by familiar and unfamiliar voices may rely on separate cognitive processes. The idea that different neural mechanisms are evoked for familiar and unfamiliar voices has recently been proposed for voice identity processing (Maguiness, Roswandowitz, & Von Kriegstein, 2018).

Alternatively, words spoken by familiar and unfamiliar talkers may both undergo the same normalization process, but this process could be more efficient for familiar than unfamiliar voices (Nygaard & Pisoni, 1998; Yonan & Sommers, 2000). One method by which this could be

realized is if listeners store information about the acoustic properties of a familiar person's voice, which they use to assist perceptual normalization. For example, listeners seem to (at least partially) rely on the vocal tract properties (formant spacing) and fundamental frequency of a familiar voice to improve intelligibility (Holmes et al., 2018). This result is consistent with theories proposing that normalization for familiar voices relies on knowledge of vocal tract properties (e.g., Peterson, 1961). For unfamiliar voices, vocal tract normalization would need to be computed online, which would be slower.

Rather than differing in their retrieval (i.e., different processes for retrieving words spoken by familiar and unfamiliar people, or different efficiencies of normalization), familiar and unfamiliar voices could instead differ in their similarity to a stored 'prototype' to which the acoustic signal is compared. The prototype theory (Lavner et al., 2001) assumes that each stimulus is compared to a representative ('central') member of the category. In the current context, this could be an acoustic stimulus that is representative of the word "shoes". Possibly, the acoustics of the familiar voice will contribute more greatly to the prototype than do either of the two unfamiliar voices, because participants have had more exposure to the familiar voice. Under this account, the prototype comparison (i.e. retrieval) process would be identical for familiar and unfamiliar talkers, but words spoken by familiar talkers would be more similar to the prototype, allowing words to be identified more rapidly. However, this explanation seems unlikely in the current context: The participants in this experiment were friends who had known each other on average for only 2.9 years, so their friend's voice would only be expected to make a small proportional contribution to their prototypes for common English words—which would include exposure to all of the voices they had encountered during their lives.

Under all of these accounts, word recognition is likely to be slower or require more cognitive resources when those words are spoken by unfamiliar talkers. Magnuson, Yamada, and Nusbaum (1995) measured reaction times for detecting target consonant-vowel sequences that were spoken by unfamiliar talkers or by the participant's family member (their spouse or

child). They found slower reaction times in blocks in which two talkers randomly alternated speaking than blocks in which one talker spoke all of the sequences. However, they found no difference in reaction times between blocks in which familiar and unfamiliar people spoke the sequences, suggesting that speech recognition is no slower for unfamiliar voices when speech is presented in isolation. Thus, instead of speeding word recognition, familiar voices may instead require fewer cognitive resources to process, which would likely not be visible in the absence of cognitive demand, such as when clear speech is presented in quiet. Nevertheless, the greater cognitive demand of recognizing words spoken by unfamiliar talkers would likely increase susceptibility to interference from linguistically similar maskers, producing a pattern of results similar to that observed here.

We found no difference in the magnitude of the familiar-voice benefit between runs in which the masking sentence was Russian or Spanish. English, Russian, and Spanish are all in the Indo-European language family, indicating common descent. Spanish is more phonetically similar to English than Russian is, which could affect the extent of informational masking (Calandruccio, Brouwer, Van Engen, Dhar, & Bradlow, 2013). Whereas, Russian is more similar to English in its prosody; Russian and English are stress-timed languages and Spanish is a syllable-timed language. Nevertheless, such differences did not appear to affect the amount of release from informational masking by a familiar voice in the foreign masker condition. Therefore, it may be the presence of speech-like information, rather than specific phonemes, that led to interference in the Russian and Spanish conditions.

When we only included talkers who acted as maskers in both the Native and Foreign conditions (i.e., the bilingual talkers), we found the same pattern of results as in the main analysis that included all talkers. Maskers in the Native condition included two bilingual takers and two unfamiliar talkers who were other participants in the experiment, whereas only two talkers were maskers in the Foreign condition. However, because each masker was presented in a separate run, there were no differences in masker variability between the conditions: In the

Native condition, we simply averaged over a greater number of adaptive runs. The magnitude of the familiar-voice benefit in the Native condition is very similar to other studies using English sentences as a masker with only two unfamiliar voices (Domingo et al., 2019; Holmes et al., 2018; Johnsrude et al., 2013).

Overall, our results demonstrate that the benefit to speech intelligibility from a naturally familiar (compared to unfamiliar) voice differs under different masking conditions. We found a large (~5 dB TMR) familiar-voice benefit to intelligibility in the presence of a same-language competing sentence, a small (~2 dB TMR) but significant benefit in the presence of an incomprehensible foreign-language sentence, and no benefit in SCN with similar spectrotemporal information as the other maskers. Participants reported more words from a target sentence correctly when it was spoken by a familiar voice, and they reported fewer incorrect words from a same-language masker sentence (with no change to the number of words reported that were spoken by neither talker). Together, these findings suggest that familiarity with a target voice improves intelligibility by helping listeners to avoid interference from distractors that are linguistically similar to the target.

Acknowledgements

This work was supported by funding from the Canadian Institutes of Health Research (CIHR; Operating Grant: MOP 133450) and the Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery Grant: 327429-2012). We would like to thank Shivaani Shanawaz, George Gainham, and Grace To for assisting with stimulus preparation and data collection.

Author Contributions

E.H. and I.S.J. designed the research. E.H. analysed the data. E.H. and I.S.J. wrote the paper.

Declaration of Conflicting Interests

The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

References

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, 45–53. https://doi.org/10.1016/j.cognition.2004.06.001

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., & Bradlow, A. R. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *American Journal of Audiology*, *22*(1), 157–164. https://doi.org/10.1044/1059-0889(2013/12-0072)

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–869. https://doi.org/10.1121/1.3458857

Darwin C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences, 1*(9), 327–333. https://doi.org/10.1016/S1364-6613(97)01097-8

Dai, B., McQueen, J. M., Hagoort, P., & Kösem, A. (2017). Pure linguistic interference during comprehension of competing speech signals. *The Journal of the Acoustical Society of America*, *141*(3), EL249-EL254. https://doi.org/10.1121/1.4977590

Domingo, Y., Holmes, E., & Johnsrude, I. S. (2019). The benefit to speech intelligibility of hearing a familiar voice. Journal of Experimental Psychology: Applied.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational

masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. https://doi.org/10.1121/1.1354984

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10615698

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16642857

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Green, T. J., & McKeown, J. D. (2001). Capture of attention in selective frequency listening. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(5), 1197–210. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11642703

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology*, *51*(7), 536–544. https://doi.org/10.3109/14992027.2012.670731

Holmes, E. (2018). Speech recording videos (Version v1.0.0) [Computer code]. Zenodo.

https://doi.org/10.5281/zenodo.1165402

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychological Science, 29*(10), 1575-1583. https://doi.org/10.1177/0956797618779083

Holmes, E., To, G., & Johnsrude, I. S. (in prep). How do voices become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Kidd, G. J., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational Masking. In W. A. Yost & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp. 143–190). Springer.

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Frontiers in Psychology*, *8*, 1584. https://doi.org/10.3389/fpsyg.2017.01584

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, *50*(3), 205–14.

Maguinness, C., Roswandowitz, C., & Von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

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).

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

Melara, R. D., Rao, A., & Tong, Y. (2002). The duality of selection: Excitatory and inhibitory processes in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 279–306. https://doi.org/10.1037//0096-1523.28.2.279

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, *35*(1), 85–103. https://doi.org/10.1016/j.wocn.2005.10.004

Norris, D., & Mcqueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

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

Perrone-Bertolotti, M., Tassin, M., & Meunier, F. (2017). Speech-in-speech perception and executive function involvement. *PLoS ONE*, *12*(7), 1–20. https://doi.org/10.1371/journal.pone.0180084

Peterson, G.E. (1961) Parameters of vowel quality. *Journal of Speech and Hearing Research 4,* 10-29.

Pollack, I. (1975). Auditory Informational Masking. *Journal of the Acoustical Society of America*, *57*, S5.

Rosen, S., Souza, P. E., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–43. https://doi.org/10.1121/1.4794379

Samson, F., & Johnsrude, I. S. (2016). Effects of a consistent target or masker voice on target speech intelligibility in two- and three-talker mixtures. *The Journal of the Acoustical Society*

of America, 139(3), 1037–1046. https://doi.org/10.1121/1.4942589

Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *The Journal of the Acoustical Society of America*, *90*(3), 1332–9. https://doi.org/10.1121/1.401925

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. J Exp Psychol Learn Mem Cogn, 13(1), 64–75.

Souza, P. E., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, *24*(January 2013), 689–700. https://doi.org/10.3766/jaaa.24.8.6

Treisman, A. (1964). The Effect of Irrelevant Material on the Efficiency of Selective Listening. *The American Journal of Psychology*, *77*(4), 533–546. https://doi.org/10.2307/1420765

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J Acoust Soc Am*, *121*(1), 519–526.

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test. *International Journal of Audiology*, *54*(sup2), 35–43. https://doi.org/10.3109/14992027.2015.1020969

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, *18*(1), 1–10. https://doi.org/10.1111/j.2044-8317.1965.tb00689.x

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*(I), 88–99. https://doi.org/10.1037/0882-7974.15.1.88

Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-noise tests for multilingual hearing screening and diagnostics. *American Journal of Audiology, 22*(1), 175–178. https://doi.org/10.1044/1059-0889(2013/12-0061

Table 1

Sub-set of the English version of the Oldenburg International Matrix corpus that were used in the experiment. Sentences with the Name word 'Peter' were used as target sentences only. Sentences with the Name words 'Kathy' and 'Rachel' were used as masker sentences in the Native condition only.

Name	Verb	Number	Adjective	Noun
Peter	got	three	large	desks
Kathy	sees	nine	small	chairs
Rachel	brought	seven	old	tables
	gives	eight	dark	toys
	sold	four	heavy	spoons
	prefers	nineteen	green	windows
	has	two	cheap	sofas
	kept	fifteen	pretty	rings
	ordered	twelve	red	flowers
	wants	sixty	white	houses

Table 2

Sub-set of the Russian version of the Oldenburg International Matrix corpus that were used in the Foreign condition of the experiment.

Name	Verb	Number	Adjective	Noun
Юрий	берёт	сто	главных	фильмов
	видит	двести	красных	часов
	даёт	мало	лучших	шаров
	делает	много	нужных	газет
	ищет	Пять	разных	залов
	купит	шесть	серых	книг
	любит	семь	старых	комнат
	найдёт	восемь	целых	марок
	помнит	девять	чужих	рядов
	хочет	десять	больших	улиц

Table 3

Sub-set of the Spanish version of the Oldenburg International Matrix corpus that were used in the Foreign condition of the experiment.

Name	Verb	Number	Adjective	Noun
Carlos	tiene	cuatro	barcos	lindos
	hace	veinte	platos	baratos
	toma	ocho	regalos	negros
	busca	mil	guantes	grandes
	quiere	dos	zapatos	viejos
	compra	tres	juegos	nuevos
	pinta	doce	dados	pequeños
	mira	siete	sillones	enormes
	pierde	seis	anillos	azules
	vende	diez	libros	bellos

Table 4

Target and masking stimuli for each adaptive run in the Speech Intelligibility task. The target stimulus was always an English sentence. SCN = Signal correlated noise.

Run ID	Familiarity condition	Masker condition	Target talker	Masking stimulus
A	Familiar	Native	Partner	English sentence: Unfamiliar 1
B	Familiar	Native	Partner	English sentence: Unfamiliar 2
C	Familiar	Native	Partner	English sentence: Russian bilingual
D	Familiar	Native	Partner	English sentence: Spanish bilingual
E	Unfamiliar	Native	Unfamiliar 1	English sentence: Unfamiliar 2
F	Unfamiliar	Native	Unfamiliar 1	English sentence: Russian bilingual
G	Unfamiliar	Native	Unfamiliar 1	English sentence: Spanish bilingual
H	Unfamiliar	Native	Unfamiliar 2	English sentence: Unfamiliar 1
I	Unfamiliar	Native	Unfamiliar 2	English sentence: Russian bilingual
J	Unfamiliar	Native	Unfamiliar 2	English sentence: Spanish bilingual
K	Familiar	Foreign	Partner	Russian sentence: Russian bilingual
L	Familiar	Foreign	Partner	Spanish sentence: Spanish bilingual
M	Unfamiliar	Foreign	Unfamiliar 1	Russian sentence: Russian bilingual
N	Unfamiliar	Foreign	Unfamiliar 1	Spanish sentence: Spanish bilingual
O	Unfamiliar	Foreign	Unfamiliar 2	Russian sentence: Russian bilingual
P	Unfamiliar	Foreign	Unfamiliar 2	Spanish sentence: Spanish bilingual
Q	Familiar	SCN	Partner	SCN from Unfamiliar 1
R	Familiar	SCN	Partner	SCN from Unfamiliar 2
S	Familiar	SCN	Partner	SCN from Russian bilingual speaking Russian
T	Familiar	SCN	Partner	SCN from Spanish bilingual speaking Spanish

U	Unfamiliar	SCN	Unfamiliar 1	SCN from Unfamiliar 2
V	Unfamiliar	SCN	Unfamiliar 1	SCN from Russian bilingual speaking Russian
W	Unfamiliar	SCN	Unfamiliar 1	SCN from Spanish bilingual speaking Spanish
X	Unfamiliar	SCN	Unfamiliar 2	SCN from Unfamiliar 1
Y	Unfamiliar	SCN	Unfamiliar 2	SCN from Russian bilingual speaking Russian
Z	Unfamiliar	SCN	Unfamiliar 2	SCN from Spanish bilingual speaking Spanish