The effects of target-masker sex mismatch on linguistic release from masking

Brittany T. Williams^{a)} and Navin Viswanathan

Department of Communication Sciences and Disorders, The Pennsylvania State University, University Park, Pennsylvania, 16801 USA

ABSTRACT:

Listeners often experience challenges understanding an interlocutor (target) in the presence of competing talkers (maskers). However, during linguistic release from masking (LRM), this difficulty decreases for native language targets (English) when paired with different language maskers (e.g., Dutch). There is considerable evidence that the linguistic similarity between target-masker pairs determines the size of LRM. This study investigated whether and how LRM is affected when the streams also differed in talker sex. Experiment 1 investigated intelligibility for English targets in sex-matched and mismatched conditions with Dutch or English maskers. While typical LRM effects were obtained when sex was matched, *opposite effects* were detected when sex was mismatched. In experiment 2, Mandarin maskers were used to increase linguistic dissimilarity and elicit stronger LRM effects. Despite the greater linguistic dissimilarity, the surprising reverse LRM effect in the sex-mismatch condition persisted. In experiment 3, the target stream was held constant and talker sex and language were manipulated in the masker. Here, expected LRM effects were obtained for both the sex-matched and sex-mismatched conditions. This indicated that the locus of the dissimilarities and not just relative properties affect LRM. Broadly, this study suggests that using naturally varying listening situations advances understanding of factors underlying LRM.

© 2020 Acoustical Society of America. <https://doi.org/10.1121/10.0002165>

(Received 29 January 2020; revised 11 September 2020; accepted 22 September 2020; published online 13 October 2020)

[Editor: Matthew J. Goupell]

Pages: 2006–2014

I. INTRODUCTION

Understanding speech successfully in environments with multiple competing speakers is a challenge for listeners. This situation has often been referred to as the cocktail party problem (Cherry, 1953). For instance, in a noisy café, a listener must successfully separate their interlocutor's speech (the target) from other speakers and sound sources present (the maskers), such as music, air conditioners, footsteps, etc. Moreover, listeners cannot entirely disregard the extraneous sounds in their environment because these, too, may be necessary for effective communication. For instance, approaching footsteps and spoken interference from waitstaff require listeners to shift attention from their current interlocutor's speech when ordering a meal. Thus, the ability to segregate speech from competing sound sources in a listening environment is essential for communicative success. How listeners successfully do so is a critical question for accounts of speech perception to address. Several studies show that speech intelligibility increases when target and masker are more dissimilar. Such increased intelligibility has been demonstrated, for instance, when the speech streams differ in speaker sex (e.g., Brungart, 2001) or spatial location (e.g., Litovsky, 2005; Freyman *et al.*, 1999). These increases in intelligibility are referred to as *release from masking* because listeners benefit from the

differences between target and masker such that speech perception is improved.

A. Linguistic release from masking

A particularly interesting example of improved intelligibility arises due to the presence of maskers in a different language from that of the target speech. In such cases, improved intelligibility for different language maskers for listeners, known as linguistic release from masking (LRM; Viswanathan *et al.*, 2016; Calandruccio *et al.*, 2016), has been observed. For example, native English listeners experience LRM when listening to English targets presented in competing Dutch maskers relative to English maskers (Freyman *et al.*, 2001; Brouwer *et al.*, 2012). LRM effects have also been demonstrated for English targets presented in Mandarin (Van Engen and Bradlow, 2007). These effects have been explained by the target-masker *linguistic similarity hypothesis* in which more linguistically dissimilar target and masker speech streams are easier to segregate (Brouwer *et al.*, 2012). Particularly, LRM increases when the linguistic distance is increased between the target and masker. Linguistic distance, here, is defined as the extent to which languages differ across lexical, syllabic, and sound level as defined by typology and language family (e.g., Shei, 2014) is increased between the target and masker.

B. Understanding factors underlying LRM

Consistent with general accounts of speech-on-speech masking, at least two possible factors may explain LRM.

^{a)}Electronic mail: btw@psu.edu, ORCID: 0000-0002-6779-8331.

First, LRM might occur because of differences in *energetic masking* in the two masker language conditions. Specifically, linguistic differences at the segmental (phonetic/phonological) and suprasegmental (stress patterns, prosody) levels of speech lead to differences in the energetic composition of the background speech, leading to a differing overlap with the target speech. For example, target-masker combinations that differ widely on a linguistic-phonetic dimension show a notable LRM. Calandruccio *et al.* (2010) used speech-shaped noise maskers instead of speech maskers to isolate the energetic characteristics of background speech. These speech-shaped noise maskers retained the spectral and temporal properties derived from the original speech maskers. LRM-like effects still occurred for the Mandarin speech-shaped noise maskers relative to English despite the absence of linguistic content. This suggests that energetic characteristics of the masker contribute to the LRM effect obtained with the original speech masker. This study also demonstrated that listeners showed improved intelligibility for the English speech-shaped noise masker relative to English speech maskers. This indicated that there were residual informational masking contributions to LRM. Broadly, this study, among others, suggests a role for energetic masking in LRM (Calandruccio *et al.*, 2013; Viswanathan *et al.*, 2016, 2018).

A second more prominent route to LRM is the differences in *informational masking* that occur because the phonological, lexical, syntactic, and semantic characteristics of the masking languages differentially interfere with the target speech. For example, Brouwer *et al.* (2012) presented English targets with English and Dutch maskers that were either semantically meaningful or anomalous to investigate informational masking by focusing on the role of semantic interference in LRM. While this study showed LRM effects in both conditions, it revealed that semantic interference between the target and masker had a significant effect on listener performance. Specifically, listeners showed greater intelligibility when presented with semantically anomalous English maskers compared to the meaningful ones. Thus, it is clear that informational factors, such as semantic properties of the masker, influence the size of LRM (cf. Calandruccio *et al.*, 2018, which indicates that differences in semantic meaningfulness do not always lead to differences in LRM). The role of informational masking effects in speech-on-speech masking have been examined more generally by testing the role of learning. Dai *et al.* (2017) used two- or four-band noise-vocoded (spectrally degraded) speech maskers to test speech-on-speech recognition. This approach allows informational content through reduced intelligibility to change while the acoustic signal itself is unchanged. During the experiment, the four-band noise-vocoded masker speech was unintelligible to listeners in the beginning, but after training on vocoded speech, listeners showed an increase in masker intelligibility, which lead to more interference perceiving the target. While Dai and colleagues did not examine LRM specifically, their findings demonstrate that linguistic content or semantically

meaningful vs anomalous maskers plays a significant role in separating target-masker pairs.

Broadly, even though the factors underlying LRM have been subject to careful study and are well-described, the relative contributions of the different factors and how their effects vary in different listening conditions are still unclear. In the current study, we focus on a different aspect of understanding LRM.

C. Understanding how LRM changes in different listening conditions

Past studies have also examined how LRM responds to changing listening conditions. In particular, testing LRM under more ecologically valid listening conditions provides complementary evidence of factors essential to LRM (Bronkhorst, 2015). Recent studies have demonstrated that manipulating factors like spatial location to modify listening conditions affects the magnitude of LRM. For instance, Viswanathan *et al.* (2016) investigated whether LRM effects persist when the maskers are presented from separate locations relative to the target (spatial separation). This work augments typical LRM studies that examine LRM effects when the target and masker originate from the same location (colocated). First, they replicated past findings by showing that listener performance was overall better in the spatially separated vs spatially colocated conditions. Interestingly, for our purposes, the researchers found typical LRM effects under altered listening conditions. That is, LRM still persisted when the masker was spatially separated from the target (i.e., easier segregability). The magnitude of LRM diminished relative to the colocated (i.e., harder segregability) condition. Similarly, Viswanathan *et al.* (2018) extended the prior investigation to ask whether LRM effects persisted under a different set of altered listening conditions. Here, they tested this using speech streams spectrally degraded via noise-vocoding. Results showed that the magnitude of LRM, again, varied when the listening situation was changed. Specifically, access to the acoustic information in the target modulated LRM effects by using noise-vocoded speech with high (easier segregability) and low (harder segregability) spectral resolution. Finally, Viswanathan and Kokkinakis (2019) demonstrated that altering listening situations through varying the extent of reverberation can also determine whether or not LRM effects are obtained with the same target-masker pairs. Overall, these studies suggest that changes in listening situations affect the size of the obtained LRM.

D. Focus of the current study: Examining simultaneous effects of multiple sources of target-masker dissimilarity

Findings from Viswanathan *et al.* (2016, 2018) and Viswanathan and Kokkinakis (2019) raise the possibility that target-masker linguistic similarity alone might not fully explain the presence and size of LRM. In the present study, we examine this possibility further by studying LRM in

conditions in which the mismatch in talker-sex between the target and masker is manipulated. That is, we will introduce an additional source of dissimilarity between the competing streams, in addition to variation along the linguistic dimension. This manipulation likely affects the informational and energetic overlap between competing streams but does not alter the nature of the background speech signal fundamentally (cf. speech-shaped noise or noise-vocoded speech manipulations described earlier). Specifically, we will use expected effects of sex mismatch between the streams in order to determine whether and how this affects LRM.

While heretofore there have not been studies of LRM using sex mismatch, several studies have investigated the effects of talker sex-mismatch in speech-on-speech situations (Festen and Plomp, 1990; Brungart, 2001; Helfer and Freyman, 2008). These studies demonstrate that when talkers differ in sex, masking effects are reduced overall (Brungart, 2001; Helfer and Freyman, 2008). For example, Brungart (2001) investigated intelligibility of targets in the presence of competing maskers when the target-masker pairs were produced by either the same talker, different talkers of the same sex, or talkers of different sexes. Brungart found that speech intelligibility was best when target-masker pairs were produced by different-sex talkers. Such findings are often attributed to minimization of informational masking effects in the different-sex talker condition relative to the same-sex talker condition (Brungart, 2001). In addition, energetic differences may also contribute to better performance in such situations. This is because of clear differences in acoustic characteristics for male and female speech related to vocal fold length, as well as shape and size of the vocal tract, including fundamental frequency and dispersion of formant frequencies (Titze, 1989; Fant, 1966).

Whether and how sex mismatch affects the resulting LRM is unknown. Thus, consistent with past studies, we anticipate that a manipulation of the sex mismatch of the competing streams will affect LRM through both the informational and energetic routes. In experiment 1, we investigated LRM effects for English targets (matched or mismatched in sex) with Dutch or English maskers. In experiment 2, we further explored these effects using Mandarin maskers that were linguistically more dissimilar to English compared to Dutch. Finally, in experiment 3, we tested whether the obtained effects persist when both factors (talker sex and language) are simultaneously manipulated in the masker.

II. EXPERIMENT 1

In this experiment, we introduced an additional source of stream dissimilarity by using target-masker pairs that are mismatched in sex to examine how LRM is modulated. To do so, we used a male and female native talker of American English as targets in the presence of competing Dutch or English female masker speech.

A. Method

1. Participants

Thirty-seven listeners (15 females and 22 males), ranging from 18 to 28 years old (median age = 19.16 yr), provided informed consent to participate in the experiment. All participants who were native speakers of American English had no knowledge of Dutch. Participants reported no hearing or speech deficits. In addition, participants were screened for previous significant exposure to foreign-accented speech and completed a demographic questionnaire.

2. Materials

One female and one male native English speaker produced the target stimuli. Following past masking studies, syntactically simple sentences from the revised Bamford-Kowal-Bench (BKB-R) Standard Sentence Test were selected for recordings (Bench *et al.*, 1979). A total of 64 sentences were chosen from the BKB-R sentence list. Each sentence contained 3 or 4 key words for a total of 256 key words (e.g., *The CLOWN had a FUNNY FACE.*). All masker stimuli were produced by two native English and two native Dutch female speakers. Masker stimuli were composed of semantically anomalous sentences (e.g., *The black top ran the spring.*) to focus on the contributions of typological properties of masking languages, such as phonetic mismatches, as well as suprasegmental characteristics. Following Brouwer *et al.* (2012), we used the syntactically normal sentence test (SNST) for masker content (Nye and Gaitenby, 1974). We obtained Dutch translations of SNST sentences from Brouwer *et al.* (2012) to elicit Dutch masker stimuli. All stimuli were recorded in a sound-attenuating booth at a 44.1 kHz sampling rate and a 16-bit resolution.

First, the root-mean-square (RMS) levels of the target sentences were equalized to the same overall level and set at 66 dB sound pressure level (SPL). Then, 500 ms of silence was added to the beginning and end of the target for a total of 1 s in order to match the duration of the masker track. After editing the target sentences, two-talker babble, consisting of a combined waveform of two speakers, was created for each background condition. Competing speech, consisting of two speakers, has shown robust informational masking effects (Brungart, 2001; Calandruccio *et al.*, 2010). The two-talker Dutch and English maskers were created using Praat (Boersma and Weenink, 2014). Following Calandruccio *et al.* (2010), all maskers were equalized to the same RMS level to produce a signal-to noise-ratio (SNR) of -5 dB when combined with targets set at 66 dB SPL. This SNR level has been reported to be challenging enough to allow for the detection of both informational and energetic masking effects (Calandruccio *et al.*, 2010; Kidd *et al.*, 2016; Kidd *et al.*, 2019). The match and mismatch conditions were counterbalanced for all participants.

The long-term average speech spectrum (LTASS) of the background speech tracks was normalized following the procedure by Brouwer *et al.* (2012) to reduce unequal

TABLE I. Brief overview of the stimuli design for experiments 1–3.

Experiment 1		Experiment 2		Experiment 3	
Target _{FEMALE}	Target _{MALE}	Target _{FEMALE}	Target _{MALE}	Target _{FEMALE}	
English masker _{FEMALE}		English masker _{FEMALE}		English masker _{FEMALE}	English masker _{MALE}
Dutch masker _{FEMALE}		Mandarin masker _{FEMALE}		Mandarin masker _{FEMALE}	Mandarin masker _{MALE}

amounts of energetic masking between background language conditions. Finally, the target and masker sentences were combined using Praat (Boersma and Weenink, 2014). For each target sentence, a random portion of the corresponding masker was selected.

3. Procedure

Listeners were seated in front of a computer and tested individually. Participants received oral instructions. They listened to English sentences spoken by a native English speaker in the presence of two-talker background speech and typed what they heard from the target speaker. They were instructed to report individual words if they were unable to identify the whole sentence. Each sentence was played once. Participants provided their response before initiating the next trial.

Prior to the experimental trials, the listener heard eight practice trials (from a BKB-R list) to familiarize themselves with the task and the target talker. The practice trials were presented at 0 dB SNR. During the experimental trials, one target sentence was played on each trial with the masker. The experimental trials were presented at −5 dB SNR. The masker was presented 500 ms prior to the beginning of the sentence and 500 ms at the end of the sentence for a total of 1 s longer than the target sentence.

The presentation order of each masker condition (English and Dutch) was randomized for all participants. As seen in Table I, listeners were presented with 16 sentences per masker language per match (2 masker languages × 2 talker sexes × 16 sentences = 64 sentences total). Participants' responses were scored based on 3–4 key words

per sentence, according to the key words specified in the BKB-R sentence lists (Bench *et al.*, 1979).

B. Results

Listener performance was measured by scoring the correct key words reported per sentence in order to obtain an intelligibility score. As expected, the leftmost panel of Fig. 1 indicated that listeners reported more key words correctly when target and masker were mismatched in sex. But surprisingly, listeners recognized fewer key words in the different language background with a strong interaction evident between masker language and talker sex-match.

The percent correct scores were transformed into logit values to normalize error variance before inferential analyses. The transformed data were submitted to a 2 (masker language: English and Dutch) × 2 (talker sex match: mismatch and match) repeated measures factorial analysis of variance (ANOVA) to examine the effect of masker language and match on intelligibility. The analysis showed that intelligibility was not significantly influenced by talker sex match alone, $F(1,36) = 2.99$, $p = 0.09$, $\eta^2_p = 0.08$, although it did reveal typical LRM effects in that Dutch maskers produced better intelligibility than English maskers overall, $F(1,36) = 12.56$, $p = 0.001$, $\eta^2_p = 0.26$. However, it is important to note that the overall typical LRM effects obtained were affected by a strong, significant interaction between talker sex match and masker language, $F(1,36) = 109.22$, $p < 0.001$, $\eta^2_p = 0.75$. As shown in Fig. 1, listeners perform better in the Dutch masker condition compared to the English masker condition but only when the

FIG. 1. The effect of match and masker language on intelligibility for experiments 1–3. The difference between the white and gray bars indicates LRM in each stream condition. The change in direction indicates the LRM in the match condition but the reverse LRM in the mismatch condition. Error bars represent standard errors.

talker sex was matched. When the talker sex was mismatched, this effect flipped, giving us reverse LRM effects.

To probe this interaction, we performed a *post hoc* analysis of simple main effects. This analysis examined whether masker language had significant effects on intelligibility for listeners at the level of sex-match followed by sex-mismatch. Results revealed that listeners showed typical LRM effects in the sex-match condition, $F(1,36)=49.75$, $p < 0.001$, $\eta_p^2=0.58$, producing more key words correct for Dutch than for English when the talkers were matched in sex. However, reverse LRM effects were shown in the sex-mismatch condition, $F(1,36)=31.44$, $p < 0.001$, $\eta_p^2=0.47$, in which listeners produced the opposite pattern. In addition to intelligibility scores, we also report LRM scores that are calculated by subtracting the intelligibility in English from the Dutch conditions in both sex-match and sex-mismatch conditions for each subject as shown in the first panel of Fig. 2.

Overall, listeners show masking release for English language maskers when talker sex is mismatched. During LRM, Dutch language maskers typically show masking release relative to English maskers in the presence of an English target. Here, we see a reversal of LRM when sex is mismatched, which contradicts the target-masker linguistic similarity hypothesis.

C. Discussion

In this experiment, listeners were presented with mismatch target-masker sex pairs to investigate how and whether LRM would be affected by the simultaneous presence of another source of stream dissimilarity. Listeners

performed better overall with a masker language different from English but not when the target and masker differed in sex. A further examination indicated that the masker language effects strongly interacted with mismatch in talker sex between the target and the masker language. Interestingly, while listeners in the match condition show the expected LRM, this pattern reverses in the mismatch condition in which listeners performed better when presented with English compared to Dutch maskers.

Critically, this experiment shows that similarity between the target and masker language alone does not automatically determine performance. Here, we uncovered a condition in which the performance was actually better with a native language masker (i.e., English) when the sex is mismatched. This pattern of results raises some questions. Do these unexpected reverse LRM effects generalize across other language maskers? Specifically, does reverse LRM only occur in conditions in which the LRM effects are relatively weak?, for example, when the target and masker are linguistically similar as in the current study? We conducted follow-up experiments to answer these questions.

III. EXPERIMENT 2

In experiment 2, we tested whether we would replicate reverse LRM using Mandarin as a masker language that is different from English. We expected strong LRM effects with Mandarin because it is significantly different from English in contrast with Dutch (Calandruccio *et al.*, 2013). Based on the target-masker similarity hypothesis (Brouwer *et al.*, 2012), we expect stronger LRM effects in experiment

FIG. 2. The LRM effect plotted in percentage points for all listeners for each sex-match or mismatch condition in experiments 1–3. Boxes below the “0” line on the y axis denote reverse LRM effects. Each dot represents the estimated benefit of LRM for every individual. The boxes depict the values between the 25th and 75th percentiles, and the whiskers denote the ± 1.5 interquartile range. Medians are shown as horizontal lines.

2 and no reversal of LRM effects when the streams were mismatched in sex.

A. Method

1. Participants

Twenty-six listeners (10 females and 16 males), ranging from 18 to 23 years old (median age = 19.12 yr), participated in the experiment. All participants were native speakers of American English with no knowledge of Mandarin and provided informed consent. Participants reported no hearing or speech deficits and provided information on exposure to foreign-accented speech.

2. Materials and procedure

All materials and procedures were the same as in experiment 1, including target speakers. The only exception was replacing Mandarin as a masker instead of Dutch (see Table I). Therefore, the Mandarin sentences were composed of the same sentences from the SNST. The sentences were translated into Mandarin and produced by two native Mandarin female speakers.

B. Results

This pattern of results closely resembled those of experiment 1. As shown in the middle panel of Fig. 1, reverse LRM does persist even with linguistically dissimilar languages, and listeners still report more key words when target and masker pairs are mismatched in sex.

Similar to experiment 1, data were submitted to a 2 (masker language, English and Mandarin) $\times$ 2 (match, mismatch and match) repeated measures factorial ANOVA to examine the effects of masker language and match on intelligibility.

Listeners reported more key words accurately when the sex of the streams was mismatched, $F(1,25) = 19.79$, $p < 0.001$, $\eta_p^2 = 0.44$. Listeners demonstrated higher intelligibility in the presence of Mandarin rather than English maskers, $F(1,25) = 15.38$, $p = 0.001$, $\eta_p^2 = 0.38$. However, this effect, too, was determined by the strong sex-match and masker language interaction, $F(1,25) = 111.64$, $p < 0.001$, $\eta_p^2 = 0.70$. This indicates that the effect of the masker language depended on whether the talker sex was mismatched (see Fig. 1).

To further explore the interaction, we performed a *post hoc* analysis of simple main effects. This analysis revealed that masker language showed significantly better performance for listeners in Dutch vs English in the sex-match condition, $F(1,25) = 49.00$, $p < 0.001$, $\eta_p^2 = 0.67$, but the opposite pattern, $F(1,25) = 26.00$, $p < 0.001$, $\eta_p^2 = 0.51$, in the sex-mismatch condition. Overall, these results provide further support for the surprising reverse LRM effect when talkers are mismatched in sex. This interaction is illustrated in the middle panel of Fig. 2 using LRM scores. Note that these scores were calculated by subtracting the intelligibility

in English from the Mandarin conditions in both sex-match and sex-mismatch conditions for each subject.

C. Discussion

Listeners performed better when the target-masker pairs were mismatched by sex and masker language as expected. However, the surprising reverse LRM effect *still occurred* in the mismatch condition. That is, listeners performed better with the English masker compared to Mandarin. Given the persistence of reverse LRM in linguistically dissimilar and similar target-masker pairs, we can assume reverse LRM is not attributable to the overall weaker LRM effects obtained with Dutch maskers.

What then could explain the reverse LRM effects in experiments 1 and 2? We speculate on two possibilities. The first possibility is that mismatch in talker sex already significantly improves the segregability of the streams as indicated by the strong effects of sex mismatch in both studies. In this case, when the streams have already been segregated, familiarity in masker content may actually be beneficial to the listeners, and they may be able to generate expectations based on prior knowledge of the unfolding speech stream (Russo and Pichora-Fuller, 2008). The second possibility is that talkers might themselves differ in their susceptibility to masking effects. To be clear, pilot testing confirmed that targets used in the first two experiments were completely intelligible in quiet. Generally, the locus of the manipulation should not matter because only the relative properties of the target and masker are considered to drive intelligibility. Under this assumption, we manipulated the sex mismatch in the target stream (female vs male target talker) and language in the masker stream (female English talkers vs female non-English talkers) in the first two experiments. However, if the two target talkers were somehow differentially affected by the presence of maskers, then the locus of sex and language mismatch could change the obtained LRM.

IV. EXPERIMENT 3

In this study, we opted to keep the target streams the same while manipulating both the language and sex mismatch in the masker streams. We did so by using the same manipulations as were used in experiment 2 with the crucial difference that the target stream held constant across all conditions.

A. Method

1. Participants

Thirty-four listeners (19 females and 15 males), ranging from 18 to 25 years old (median age = 18.97 yr), participated in the experiment. All participants were native speakers of American English with no knowledge of Mandarin and provided informed consent. Participants reported no hearing or speech deficits. In addition, participants were screened for previous significant exposure to foreign-accented speech and completed a demographic questionnaire.

2. Materials and procedure

All materials were the same as in experiment 2 except that the sex of the target speakers remained constant (female), whereas the sex of the maskers was either mismatched (female target + male masker) or matched (female target + female masker). All targets were English, whereas the maskers were either English or Mandarin. The procedure was similar to the previous experiments (see Table I). However, the main experimental task was not blocked by the target sex because there was only one target speaker.

B. Results

Listener performance was again measured by scoring the correct key words reported per sentence in order to obtain an intelligibility score. The right panel of Fig. 1 reveals that the typical LRM effects return when the masker language and talkers mismatched in sex were manipulated in the masker stream. As expected, listeners reported more key words in the sex-mismatch condition compared to the sex-match condition.

As with experiments 1 and 2, logit-transformed data were submitted to a 2 (masker language, English and Mandarin) $\times$ 2 (talker sex-match, mismatch and match) repeated measures factorial ANOVA to examine the effects of masker language and match on intelligibility. Listener performance is shown in Fig. 1. The analysis revealed the typical LRM effect for masker language such that overall listener performance was better for Mandarin maskers than for English maskers, $F(1,29) = 50.20$, $p < 0.001$, $\eta_p^2 = 0.63$. In addition, overall intelligibility was better for sex-mismatched conditions than for sex-matched conditions, $F(1,29) = 180.80$, $p < 0.001$, $\eta_p^2 = 0.86$. This time, the analysis did not reveal a significant interaction between sex-match and masker language, $F(1,29) = 1.09$, $p = 0.304$, $\eta_p^2 = 0.036$. This indicated that overall performance was better for Mandarin across match and mismatch conditions, resulting in no reverse LRM effect and only typical LRM effects (see Fig. 1). These results are further illustrated in the last panel of Fig. 2 by LRM scores calculated by subtracting the intelligibility in English from the Mandarin conditions in both the sex-match and mismatch conditions for each subject.

C. Discussion

In experiment 3, we obtained expected effects of sex and language dissimilarities but no evidence of reverse LRM when the talker sex was mismatched between streams. This demonstrates that listeners benefit both when maskers differ in language as well as when maskers differ in talker sex. In particular, unlike in the two previous experiments, typical LRM effects occurred even when there was a mismatch in talker sex. But this pattern arises only when talker sex is manipulated for the masker not in the target. Given the results of this experiment, it is clear that the locus of differences between sex-mismatched talkers influences LRM

effects more than the overall relative properties of talkers alone.

V. GENERAL DISCUSSION

The present study investigated how factors in a listeners' acoustic environment affect LRM. We examined LRM effects in the presence of talker sex-mismatch. Particularly, we investigated the influence of sex-mismatch on speech intelligibility of a native language (i.e., English) in the presence of a masker language different from the target (i.e., Dutch or Mandarin). Specifically, adding an additional source of stream dissimilarity greatly influenced LRM. In line with previous studies, LRM effects differed depending on the listening situation. In the mismatch condition for experiments 1 and 2, we obtained LRM effects in the opposite direction (reverse LRM) when we manipulated talker sex in the target and not the masker. However, in experiment 3, LRM effects in the expected direction were found when we manipulated the talker sex in the masker.

In the rest of the general discussion, we outline three major conclusions drawn from this set of experiments.

A. Target-masker linguistic similarity does not completely predict LRM

The performances in the mismatch conditions of experiments 1–3 are not predicted by linguistic similarity alone. Given that the sex-match condition shows the same pattern across experiments, it is reasonable that earlier accounts of LRM focused on language similarity as the primary explanation. However, with the introduction of the simultaneous mismatch in talker sex, an interesting mix of resulting patterns was obtained. LRM effects were modulated by the locus of linguistic differences in the competing streams rather than just overall dissimilarity between streams. Specifically, both LRM (experiment 3) and reverse LRM effects (experiments 1 and 2) were obtained when the target and masker were mismatched in sex and the effects also depended on the locus of our manipulation. Taken together, our findings indicate that when multiple sources of stream dissimilarity exist, linguistic similarity, alone, is insufficient to predict the size and direction of the resulting LRM.

B. What underlies reverse LRM is yet unclear

Experiments 1 and 2 were the first to detect reverse LRM effects. Through the sequence of experiments, we considered some possible explanations for reverse LRM. First, we considered that perhaps increased source uncertainty (Freyman *et al.*, 2007; Durlach *et al.*, 2003b; Brouwer and Bradlow, 2014) or difficulties in auditory object selection (Shinn-Cunningham, 2008; Ihlefeld and Shinn-Cunningham, 2008) for the matched relative to the mismatched talker sex conditions underlie these surprising reverse LRM effects. These factors have been shown to affect the relative strength of informational masking in related studies (e.g., Durlach *et al.*, 2003b; Brouwer and Bradlow, 2014). However, we obtained preliminary evidence against these explanations in

experiment 3, in which we found typical, rather than reverse, LRM effects even though we manipulated the talker sex-mismatch. The key difference was that both masker language and talker sex match or mismatch were manipulated in the background stream.

Given these results, we ponder whether individual talkers may differ in how susceptible they are to masking effects of background streams. Indeed, there is evidence in perception that groups of listeners might differ in their susceptibility to informational masking of the same signals (Kidd *et al.*, 2019; Kidd *et al.*, 2016). Perhaps there are similar differences in susceptibility to masking in production that occur due to idiosyncrasies in speaking style, tendency to hyperarticulate, characteristics of their long-term speech spectra, etc., that are not captured by measuring their intelligibility in quiet. Pursuing this line of reasoning, in future research, we intend to evaluate what factors determine talker susceptibility to different forms of masking and how it can be best characterized during LRM.

We strongly suspect that the source of this talker susceptibility might not be solely attributable to energetic factors because we could not replicate the pattern of results from experiments 1 and 2 using speech-shaped noise analogues of English, Dutch, and Mandarin maskers. Our previous study followed the same procedure as the current study and aimed to examine whether reverse LRM was mostly attributable to energetic contributions (Williams and Viswanathan, 2017; also see Calandruccio *et al.*, 2013).¹

In sum, systematic future investigation is required to both replicate reverse LRM effects, as well as pinpoint the factors that underlie these effects and the conditions under which they are obtained.

C. Relative contributions of informational and energetic masking to LRM may depend on listening conditions

Finally, the route toward clearly dissociating energetic and informational masking (in LRM and more generally) is admittedly unclear because of the unconstrained definition of informational masking (e.g., Durlach *et al.*, 2003a; Durlach, 2006; Kidd *et al.*, 2008; Mattys *et al.*, 2009; Kidd and Colburn, 2017), as well as the strong interdependency between the two types of masking. Indeed, informational masking has been used as a “catch all” for difficulty in speech-on-speech recognition that is not attributable to energetic masking alone. Given this, it is hard to determine how these components contribute to LRM in a variety of listening conditions. Inclusion of another source of stream dissimilarity, such as in the current study, likely alters their relative contributions. For instance, we know that when target and masker streams are mismatched, informational masking effects are reduced (Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Brouwer *et al.*, 2012). This reduction also occurs when the streams are specifically mismatched in sex (Brungart, 2001). However, our surprising findings appear to indicate that when the sex dissimilarity is overlaid on linguistic dissimilarity, informational

consequences are not purely additive. That is, the best performance was not obtained when both language and sex were mismatched in experiments 1 and 2. Instead, these sources of dissimilarity appear to modulate each other, making it more difficult to tell apart relative contributions. These findings further indicate the need to study this phenomenon in different listening situations in order to fully understand the challenges listeners face in typical auditory environments.

VI. CONCLUSION

Heretofore, mismatch in talker sex has not been used to explore LRM. This investigation has demonstrated that opposite effects to the typically obtained LRM (i.e., reverse LRM) might be found under certain conditions. More broadly, our studies demonstrate that linguistic similarity between the target and masker alone does not predict the existence or direction of LRM across listening conditions. Further research is required to investigate the specific conditions and the factors that produce reverse LRM effects. In conclusion, in speech-on-speech situations, how listeners use linguistic dissimilarity for perceptual segregation is constrained by other sources of similarities between the competing streams.

ACKNOWLEDGMENTS

This research was supported by a National Institutes of Health (NIH)-National Institute on Deafness and Other Communication Disorders (NIDCD) funded predoctoral training fellowship to B.T.W. (Grant No. T32 DC000052). We thank Maya Mills and Lauren Spalding for their assistance with testing the participants. We also thank Dr. Jacqueline Mogle for her guidance on *post hoc* data analysis. We are grateful to Dr. Susanne Brouwer for providing the translated SNST sentence lists and Dr. Lauren Calandruccio for providing sentence lists used in the past experiments.

¹For a link to figure results for Williams and Viswanathan (2017), visit <https://github.com/brittwilli/Masking-Projects/blob/master/SSNmde.pdf> (Last viewed October 5, 2020).

- Bench, J., Kowal, Å., and Bamford, J. (1979). “The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children,” *Br. J. Audiol.* **13**(3), 108–112.
- Boersma, P., and Weenink, D. (2014). “Praat: Doing phonetics by computer (version 5.2.16) [computer program],” available at <http://www.praat.org/> (Last viewed 31 July 2014).
- Bronkhorst, A. W. (2015). “The cocktail-party problem revisited: Early processing and selection of multi-talker speech,” *Atten. Percept. Psychophys.* **77**(5), 1465–1487.
- Brouwer, S., and Bradlow, A. R. (2014). “Contextual variability during speech-in-speech recognition,” *J. Acoust. Soc. Am.* **136**(1), EL26–EL32.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (2012). “Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content,” *J. Acoust. Soc. Am.* **131**(2), 1449–1464.
- Brungart, D. S. (2001). “Informational and energetic masking effects in the perception of two simultaneous talkers,” *J. Acoust. Soc. Am.* **109**(3), 1101–1109.

- Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., and Bradlow, A. R. (2013). "Masking release due to linguistic and phonetic dissimilarity between the target and masker speech," *Am. J. Audiol.* **22**(1), 157–164.
- Calandruccio, L., Buss, E., Bencheck, P., and Jett, B. (2018). "Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition?," *J. Acoust. Soc. Am.* **144**(6), 3289–3302.
- Calandruccio, L., Dhar, S., and Bradlow, A. R. (2010). "Speech-on-speech masking with variable access to the linguistic content of the masker speech," *J. Acoust. Soc. Am.* **128**(2), 860–869.
- Calandruccio, L., Leibold, L. J., and Buss, E. (2016). "Linguistic masking release in school-age children and adults," *Am. J. Audiol.* **25**(1), 34–40.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**(5), 975–979.
- Dai, B., McQueen, J. M., Hagoort, P., and Kösem, A. (2017). "Pure linguistic interference during comprehension of competing speech signals," *J. Acoust. Soc. Am.* **141**(3), EL249–EL254.
- Durlach, N. (2006). "Auditory masking: Need for improved conceptual structure," *J. Acoust. Soc. Am.* **120**(4), 1787–1790.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003a). "Note on informational masking (L)," *J. Acoust. Soc. Am.* **113**(6), 2984–2987.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (2003b). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," *J. Acoust. Soc. Am.* **114**(1), 368–379.
- Fant, G. (1966). "A note on vocal tract size factors and non-uniform F-pattern scalings," *Speech Transm. Lab. Q. Prog. Status Rep.* **7**(4), 22–30.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**(4), 1725–1736.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**(5), 2112–2122.
- Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (2007). "Variability and uncertainty in masking by competing speech," *J. Acoust. Soc. Am.* **121**(2), 1040–1046.
- Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.* **106**(6), 3578–3588.
- Garcia Lecumberri, M., and Cooke, M. (2006). "Effect of masker type on native and non-native consonant perception in noise," *J. Acoust. Soc. Am.* **119**(4), 2445–2454.
- Helfer, K. S., and Freyman, R. L. (2008). "Aging and speech-on-speech masking," *Ear Hear.* **29**(1), 87–98.
- Ihlefeld, A., and Shinn-Cunningham, B. (2008). "Spatial release from energetic and informational masking in a selective speech identification task," *J. Acoust. Soc. Am.* **123**(6), 4369–4379.
- Kidd, G., Jr., and Colburn, H. S. (2017). "Informational masking in speech recognition," *Aud. Syst. Cocktail Party* **60**, 75–109.
- Kidd, G., Jr., Mason, C. R., Best, V., Roverud, E., Swaminathan, J., Jennings, T., and Colburn, H. S. (2019). "Determining the energetic and informational components of speech on-speech masking in listeners with sensorineural hearing loss," *J. Acoust. Soc. Am.* **145**(1), 440–457.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources* (Springer, Boston), pp. 143–189.
- Kidd, G., Jr., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., and Best, V. (2016). "Determining the energetic and informational components of speech-on-speech masking," *J. Acoust. Soc. Am.* **140**(1), 132–144.
- Litovsky, R. Y. (2005). "Speech intelligibility and spatial release from masking in young children," *J. Acoust. Soc. Am.* **117**(5), 3091–3099.
- Mattys, S. L., Brooks, J., and Cooke, M. (2009). "Recognizing speech under a processing load: Dissociating energetic from informational factors," *Cognit. Psychol.* **59**(3), 203–243.
- Nye, P. W., and Gaitenby, J. H. (1974). "The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences," *Haskins Lab. Status Rep. Speech Res.* **37**(38), 169–190.
- Russo, F. A., and Pichora-Fuller, M. K. (2008). "Tune in or tune out: Age-related differences in listening to speech in music," *Ear Hear.* **29**(5), 746–760.
- Shei, C. (2014). *Understanding the Chinese language: A Comprehensive Linguistic Introduction* (Routledge, London and New York).
- Shinn-Cunningham, B. G. (2008). "Object-based auditory and visual attention," *Trends Cognit. Sci.* **12**(5), 182–186.
- Titze, I. R. (1989). "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.* **85**(4), 1699–1707.
- Van Engen, K. J., and Bradlow, A. R. (2007). "Sentence recognition in native-and foreign language multi-talker background noise," *J. Acoust. Soc. Am.* **121**(1), 519–526.
- Viswanathan, N., and Kokkinakis, K. (2019). "Listening benefits in speech-in-speech recognition are altered under reverberant conditions," *J. Acoust. Soc. Am.* **145**(5), EL348–EL353.
- Viswanathan, N., Kokkinakis, K., and Williams, B. T. (2016). "Spatially separating language masker from target results in spatial and linguistic masking release," *J. Acoust. Soc. Am.* **140**(6), EL465–EL470.
- Viswanathan, N., Kokkinakis, K., and Williams, B. T. (2018). "Listeners experience linguistic masking release in noise-vocoded speech-in-speech recognition," *J. Speech, Lang. Hear. Res.* **61**(2), 428–435.
- Williams, B. T., and Viswanathan, N. (2017). "Target-masker sex mismatch affects linguistic release from masking," poster presented at the *58th Annual Meeting of the Psychonomic Society*, Vancouver, British Columbia, Canada.