Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition?

Lauren Calandruccio, Emily Buss, Penelope Bencheck, et al.

Citation: The Journal of the Acoustical Society of America **144**, 3289 (2018); doi: 10.1121/1.5081679

View online: https://doi.org/10.1121/1.5081679

View Table of Contents: https://asa.scitation.org/toc/jas/144/6

Published by the Acoustical Society of America

JASA
THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA
Special Issue:
Additive Manufacturing and Acoustics
Read Now!

Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition?

Lauren Calandruccio,[1,a)] Emily Buss,[2] Penelope Bencheck,[3] and Brandi Jett[1]
[1]*Department of Psychological Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA*
[2]*Department of Head/Neck Surgery and Otolaryngology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA*
[3]*Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA*

(Received 25 January 2018; revised 7 November 2018; accepted 9 November 2018; published online 12 December 2018)

Speech-on-speech recognition differs substantially across stimuli, but it is unclear what role linguistic features of the masker play in this variability. The linguistic similarity hypothesis suggests similarity between sentence-level semantic content of the target and masker speech increases masking. Sentence recognition in a two-talker masker was evaluated with respect to semantic content and syntactic structure of the masker (experiment 1) and linguistic similarity of the target and masker (experiment 2). Target and masker sentences were semantically meaningful or anomalous. Masker syntax was varied or the same across sentences. When other linguistic features of the masker were controlled, variability in syntactic structure across masker tokens was only relevant when the masker was played continuously (as opposed to gated); when played continuously, sentence-recognition thresholds were poorer with variable than consistent masker syntax, but this effect was small (0.5 dB). When the syntactic structure of the masker was held constant, semantic meaningfulness of the masker did not increase masking, and at times performance was better for the meaningful than the anomalous masker. These data indicate that *sentence-level semantic content* of the masker speech does not influence speech-on-speech masking. Further, no evidence that similarities between target/masker sentence-level semantic content increases masking was found.
https://doi.org/10.1121/1.5081679

[VB] Pages: 3289–3302

I. INTRODUCTION

It is both intuitive and well-documented that communication for most listeners is less difficult in quiet environments compared to noisy ones. However, once noise is introduced into the environment, some groups of listeners are more detrimentally affected than others. Listeners who are particularly susceptible to masking include those with hearing loss (e.g., Plomp, 1986), older adults (e.g., Helfer and Wilber, 1990), children (e.g., Elliot *et al.*, 1979; Hall *et al.*, 2002), non-native speakers of the target language (e.g., Mayo *et al.*, 1997), and those with more limited working memory capacities (e.g., Gordon-Salant and Cole, 2016). Group differences in susceptibility to masking tend to be larger for speech maskers than noise maskers (Hall *et al.*, 2002; Calandruccio *et al.*, 2014). Given the prevalence of speech maskers in natural environments, it has been proposed that testing listeners in competing speech consisting of a small number of talkers may provide clinicians with a better indication of listeners' communication difficulties compared to speech scores in quiet or noise (Carhart and Tillman, 1970; Hillock-Dunn *et al.*, 2015; Jakien *et al.*, 2017). While this view dates back to the 1960s (Jerger *et al.*, 1968), standard clinical practice still does not include routine evaluation of masked-speech recognition with a small number of background talkers.

One potential barrier to, and likely reason for a slow implementation of clinical speech-on-speech testing, is the significant variability across stimuli used throughout different experiments. Indeed, some combinations of target/masker stimuli are significantly more challenging than others, even among same-sex talkers with comparable average fundamental frequencies (Freyman *et al.*, 2007). Generally, variability across speech stimuli can be broken down into two categories: acoustic features (e.g., voice pitch, timbre, and speaking rate) and linguistic features of the speech (e.g., semantics, syntactic structure, and lexical characteristics such as word frequency and density of phonological neighbors). However, the specific nuances of the speech stimuli that cause one target and masker pair to be significantly more challenging than another are not well understood. The goal of this report is to improve our understanding of the linguistic factors affecting speech-on-speech recognition, specifically the sentence-level semantic meaning and the syntactic structure of the masker speech.

One reason that speech-on-speech testing may be more indicative of real-world challenges in noisy environments, and therefore potentially beneficial to clinical assessment, is that this type of listening scenario causes both energetic and informational masking (Carlile, 2014). Energetic masking describes the degradation of signal encoding in the auditory periphery as a consequence of peripheral responses to the masker (Culling and Stone, 2017). In contrast, informational masking describes degradation in the central representation

[a)] Electronic mail: lauren.calandruccio@case.edu of the signal, due to a limited ability to segregate the target from the masker, and selectively attend to the signal (e.g., Billig *et al.*, 2013; Bregman, 1990; Kidd and Colburn, 2017). Informational masking is thought to stem from two distinct factors: perceptual similarity of the target and masker, and uncertainty regarding perceptual features of the signal and masker (Durlach *et al.*, 2003). Both similarity and uncertainty can be affected by a combination of acoustic and linguistic factors of the speech stimuli.

While there are numerous examples of acoustic similarity between the target and masker affecting informational masking (e.g., Calandruccio *et al.*, 2013; Festen and Plomp, 1990; Helfer and Freyman, 2008), it is less clear how similarity of linguistic features affects performance (Hoen *et al.*, 2007). If the similarity principle applies to linguistic features, one would predict that high degrees of linguistic similarity between the target and masker would cause poorer sentence recognition performance, while scenarios with less similarity between the target and masker would allow for better recognition (e.g., semantically meaningful target and semantically anomalous masker, or vice versa). In 2012, Brouwer and colleagues set out to test the *linguistic similarity hypothesis* using two different speech maskers spoken by the same two female voices. In one masker, the two females spoke meaningful sentences, and in the other masker, they spoke semantically anomalous sentences. For both masker conditions, real words and proper English syntactic structures were used. Their data indicated that for semantically meaningful target sentences the semantically meaningful masker (most linguistically similar to the target speech) was most effective, while the semantically anomalous masker (most linguistically dissimilar to the target speech) was least effective, allowing for better sentence recognition.

A close review of the methods used by Brouwer *et al.* (2012) indicates that several features of the speech used to create the masker stimuli varied between the meaningful and anomalous masker conditions in addition to the semantic meaning. For example, notable differences in syntactic structure and syllable count within words existed between the corpora used to create the two masker conditions. The semantically anomalous masker consisted of a string of sentences with identical syntactic structure and only monosyllabic words, while the semantically meaningful masker consisted of sentences that varied in syntactic structure and included both monosyllabic and disyllabic words. These differences in syntactic structure and syllable count resulted in rhythmic differences between the two maskers with the former having a more galloping rhythmic structure than the latter. Viewed in this way, variable syntactic structures within the meaningful masker condition could be viewed as a source of stimulus uncertainty, which in turn could result in more informational masking (McDermott *et al.*, 2011). Therefore, it is not clear whether the results observed in Brouwer *et al.* (2012) were due to the *linguistic similarity in semantic meaning* between the target and masker speech, or if they were due to differences in stimulus uncertainty associated with syntactic structure within the masker speech.

The purpose of the present experiments was to assess the importance of the sentence-level semantic content of the masker speech for a sentence recognition task with and without syntactic regularity. In experiment 1, the importance of the semantic content and syntactic regularity of the masker speech with respect to masked-sentence recognition was assessed while controlling for lexical items within the masker speech and the speaking rate of the speech maskers. In experiment 2, we attempted to partially replicate the Brouwer *et al.* (2012) experiment using new recordings, re-exploring the linguistic similarity hypothesis with respect to the similarity of semantic content between the target and masker speech.

A. Experiment 1: Impact of semantic content and syntactic structure of the masker speech

The first question explored in experiment 1 was the importance of semantic meaning, regardless of syntactic structure within the masker speech. Specifically, we asked whether semantic meaning changes the effectiveness of the two-talker masker when syntactic structure, vocabulary, and talker are held constant. Based on previous data, we hypothesized that a semantically meaningful masker would be more effective than a semantically anomalous masker (Brouwer *et al.*, 2012; Dai *et al.*, 2017). The second question explored in experiment 1 was the idea that differences in syntax structures could affect the rhythm and/or the predictability of the masker and, in turn, have an impact on speech-in-speech masking. That is, we compared sentence recognition with maskers that consisted of sentences that varied in syntax structure and maskers that consisted of sentences with a fixed (or predictable) syntactic structure. We hypothesized that maskers consisting of sentences with varied syntactic structure would increase uncertainty due to variable syllabic rhythm, and therefore would result in greater informational masking. A goal within experiment 1 was to minimize the acoustical differences between the competing speech maskers while varying these two sentence-level factors (meaning and syntax). To do this, maskers were created using the same talker voices and the same words [100% overlap between the meaningful and anomalous versions of the maskers that shared the same syntax (SS) structure]. These methods were used to minimize differences in the long-term magnitude spectrum and differences in speaking rate between masker conditions.

II. METHODS

A. Participants

Two groups of young-adult listeners participated in experiment 1 [group 1: $n = 20$, 16 females, 4 males, ages 18–23 yr, *mean* age = 20.4 yr, standard deviation (SD) = 1.0 yr; group 2: $n = 19$, 15 females, 4 males, ages 18–23 yr, *mean* age = 20.4 yr, SD = 1.5 yr]. All participants spoke English (American dialect) as their native language as confirmed by a completed linguistic and demographic history questionnaire form. Hearing evaluations were performed using standard clinical procedures [American Speech-Language-Hearing Association (ASHA), 2005]. All participants had hearing thresholds within the normal range of hearing [<25 dB Hearing Level (HL) bilaterally for all octave frequencies between 250 and 8000 Hz] and had clear ear canals as visualized through otoscopy. All methods used were approved by the Institutional Review Board at Case Western Reserve University. All participants were paid for their participation.

B. Stimulus development

1. Target stimuli

Target stimuli consisted of Basic English Lexicon (BEL) sentences spoken by female talker SR (see Calandruccio and Smiljanić, 2012). These sentences have simple declarative syntactic structures that vary across tokens. BEL sentences were originally created using a lexicon of speech recorded from non-native speakers of English. The lexicon was developed by interviewing and having short conversations on a wide range of topics (e.g., holidays, shopping, etc.) with 100 people that spoke English as their second language. Two examples of BEL sentences (with keywords in capital letters) are: The WORKER HURT his LEFT HAND, My GRANDFATHER MADE WOODEN CHAIRS. Thirteen BEL sentence lists were used for testing, including lists 6, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18, 19, and 20. There are 25 sentences included in each BEL list and four keywords per sentence (100 total keywords/list). These lists were chosen based on data using SR's speech productions, indicating similar mean performance between these lists. It should be noted that in the original Calandruccio and Smiljanić (2012) paper, list equivalency was determined using a speech-shaped noise masker. There are not published data establishing the list equivalencies for any speech materials in a two-talker babble that we are aware of. List equivalency could be affected when using different maskers. The average speaking rate and average fundamental frequency ($f0$) for these recordings was 3.97 syllables/s and 202 Hz, respectively.

2. Masker stimuli

Four two-talker masker conditions were created for experiment 1 [*varied syntax/meaningful* (VS/M); *varied syntax/anomalous* (VS/A); *same syntax/meaningful* (SS/M); *same syntax/anomalous* (SS/A)]. All four maskers were spoken by the same two female talkers, talker A and talker B. Talker A was a white female from upstate New York, age 23 years old. Talker B was a white female from southwestern Ohio, age 22 years old. Talkers A and B had a mean $f0 = 243$ and 193 Hz, respectively. Recordings were made in a double-walled sound-attenuated room using a KSM-42 Shure omnidirectional microphone (Niles, IL). A screen was attached to the face of the diaphragm to eliminate popping noises within the recordings. Talkers stood approximately 12 in. directly in front of the microphone. All stimuli were recorded at a 44.1 kHz sampling rate with 16-bit resolution. Neither talker had a detectable regional accent per informal judgments of five native-English speakers.

All four two-talker maskers consisted of talkers A and B speaking 20 sentences each. These 20 sentences were concatenated without pauses between sentences. Maskers differed with respect to semantic meaning (*meaningful* vs *anomalous*) and with respect to syntactic structure (*same* vs *varied syntax*). In all cases, masker speech was based on 40 sentences from the AzBio sentence corpus (Spahr *et al.*, 2012). The rationale for using this corpus is that it contains sentences that vary widely in syntactic structure and word length, while the lexical content is sufficiently diverse to support the construction of syntactically regular and/or semantically anomalous analogues. The 40 sentences we chose varied in length from 4 words to 12 words (mean length = 7.4 words/sentence) and all used a unique syntactic structure; that is, there was no repetition in syntactic structure across the sentences (see column E of supplemental material 1[1] for a list of the syntactic structures). Sixteen of the 40 original AzBio sentences were modified to replace lexically infrequent words with words that had a higher lexical frequency based on the online CELEX lexical frequency calculator (Davis, 2005; Baayen *et al.*, 1995; see supplemental material 1)[1]. The rationale for replacing infrequent with frequent words was that the resulting pool of words can be more easily rearranged to form meaningful sentences with fixed syntactic structure.

The two-talker masker for the VS/M condition consisted of the 40 sentences from the modified AzBio sentence corpus. Talkers A and B each spoke 20 of the 40 sentences. The duration of each sentence ranged from 1.2 to 3.2 s (mean duration = 2.0 s). An example of a sentence within the VS/M masker was "The boy drank the milk."

The two-talker masker for the SS/M condition was created using the lexicon of the first meaningful masker condition (VS/M). All 40 sentences for this condition shared the identical syntactic structure (i.e., there was *no* variability within syntactic structure for any of the concatenated sentences within the masker). This syntactic structure for the 40 sentences was DNVAN (D = determiner, N = noun, V = verb, A = adjective). To create these sentences, only words from the VS/M masker were used. All words were coded by type (D, N, V, or A). These words were then intentionally inserted within the DNVAN syntactic structure to create *new meaningful* sentences that did not vary in syntactic structure. The duration of each sentence ranged from 1.4 to 2.4 s (mean duration = 1.9 s). An example sentence from the SS/M masker was "The kids have dirty hair."

One consequence of manipulating masker syntax in the SS/M condition is that the number of words spoken by each masker talker differed for the SS/M and VS/M conditions. For the SS/M masker condition, there were a total of 200 words (40 total sentences × 5 words/sentence), 164 of which were unique (82%). For the VS/M masker condition there were 289 total words spoken across the 40 sentences, 186 of which were unique (64%). Despite these differences in the number of unique words between the VS and SS maskers, lexical differences were minimized by only using words from the VS/M masker condition to create the SS/M masker.

Two-talker maskers for the VS/A and SS/A conditions were based closely on associated meaningful maskers (see supplemental material 2[1]). All of the words in the VS/M and SS/M masker conditions were labeled by word type (e.g., D,

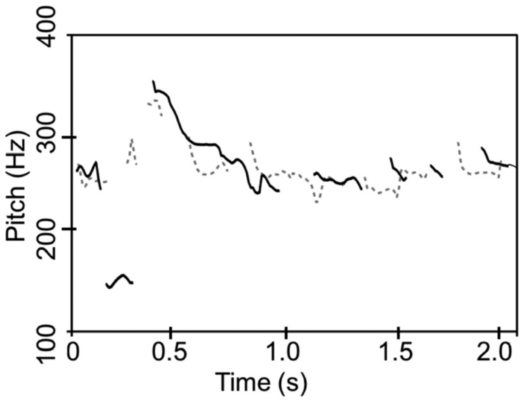

FIG. 1. Pitch contour example of two varied syntax sentences that shared the same syllable and syntax structure for one meaningful (solid black line) and one anomalous (dashed grey) sentence. As shown in this example, the pitch contours were similar between sentence pairs.

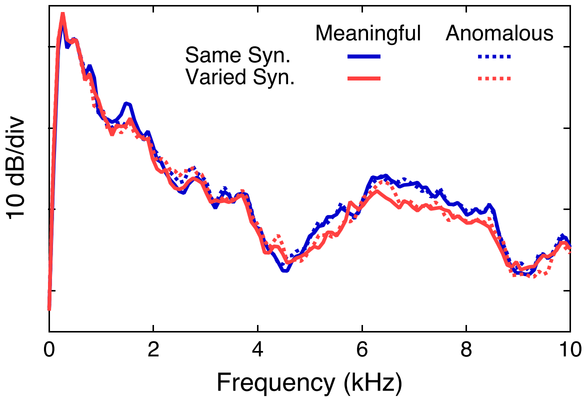

FIG. 2. (Color online) The long-term average magnitude spectrum is shown for the four two-talker masker conditions used in experiment 1. There are minimal differences between the four masker conditions and even smaller differences between pairs of maskers that share the same syntactic structure.

N, A). Words of the same type were randomly interchanged and inserted back into the syntactic structure by word type, with the caveat that the number of syllables was preserved. For example, the word type for the VS/M sentence "The little girl hurt the black cat" was defined as: DANVDAN. The corresponding VS/A sentence was "The open cards take the great plains." In this example, all words were replaced by monosyllabic words that shared the same original grammatical type of word (e.g., D, N, A) except for the two-syllable adjective "little," which was replaced by the two-syllable adjective "open," allowing for the preservation of syllable structure and word type. For the sentences included in the SS/A masker, all sentences followed the DNVAN syntactic structure. An example sentence from the SS/A masker was "The girl lives purple teens."

In an attempt to control for differences in speaking rate and prosody between the meaningful and anomalous sentence for each syntax type, talkers A and B recorded a meaningful sentence immediately followed by an anomalous sentence. For the varied syntax maskers, pairs of meaningful and anomalous sentences that shared the identical syntax were recorded sequentially. The two talkers stood in front of a computer monitor placed on the opposite side of the window in the sound-treated recording room. Talkers were presented with two sentences at a time with the meaningful sentence always displayed at the top of the screen. The words within the two sentences were aligned on the screen, and the talkers were instructed to produce the meaningful sentence first and then attempt to produce the anomalous sentence with the same cadence. These recording methods ensured that the prosody and speaking rate was similar between the meaningful and anomalous pairs (see Fig. 1). Two-tailed t-tests using an $\alpha = 0.05$ indicated no significant differences in speaking rate based on semantic meaning when the talker and syntactic structure was held constant (p-values ranging between 0.10 and 0.88); that is, there was no significant difference in speaking rate whether the talker was producing meaningful or anomalous sentences for each syntax condition. Using the same two voices for all four two-talker masker conditions helped to minimize differences in long-term average spectral components across the different masker conditions (see Fig. 2).

Recordings for the two-talker masker streams were created using the same procedures for each of the four masker conditions. Individual sentence recordings were digitally edited to remove all silences from the beginning and the end. The 20 sentences spoken by each talker were then root-mean-square (RMS) equalized and concatenated to create one continuous speech stream. The speech stream from talker A was then mixed together with the speech stream from talker B. The final few seconds of the longer of the two streams were trimmed so that the duration matched that of the shorter speech stream.

C. Procedure

Listeners were seated in a double-walled, sound-attenuated suite (Acoustic Systems, Cedar Park, TX) facing an observation window. Before testing began, listeners were told that they would be hearing three female voices played through headphones. One target sentence would be played at a time, and the task was to repeat back the sentence the target talker spoke, while trying to ignore the other two talkers. The listeners were instructed to attend to the target talker, and that the same target talker would be used throughout the duration of the experiment. Listeners were told that the target voice would start out as the loudest of the three voices, but that the intensity of the target talker would decrease as the experiment continued.

The experiment was run by a tester seated in the control room of the sound-isolated suite. Stimuli were presented to the listener using custom software developed in Max (Cycling'74, Walnut, CA) software, passed to an M-Audio Fast Track Pro (Cumberland, RI) soundcard, and output diotically through Etymōtic (Elk Grove Village, IL) ER1 insert earphones.

Listeners were instructed to repeat back exactly what they heard the target talker say, and were encouraged to do so even if they needed to guess, if they needed to guess, heard only one or a few words, or if what they heard did not make sense. Listeners' responses were scored on-the-fly by an examiner naive to the experimental hypothesis and seated in the control room of the sound suite. Responses were also recorded for later reliability scoring. Examiners were instructed to score each word exactly as they heard the participant say it, and any change to the exact keyword was scored as incorrect (e.g., deletion of pluralization or morphological ending change).

Stimuli were presented in two distinct ways. Listeners in group 1 were presented with a randomly selected segment of the masker, gated on and off on each trial. The masker was gated on 500 ms before the beginning of the target sentence and gated off 500 ms after the end of the target. The next trial began after the listener finished repeating back what they heard or indicating that they were unable to identify any of the target speech. Listeners in group 2 heard a continuous masker throughout each block of trials. In this condition, a green attention light was presented on a computer monitor placed directly in front of the listeners. On each trial a green light would turn on 500 ms before the target sentence. This was done to help ensure that the listener realized a target occurred. Without an indicator light, if a sentence was completely inaudible, then the listener would be unable to indicate to the examiner that they should proceed with the next trial. If the listener saw the green light and did not hear the target speech, they were instructed to tell the experimenter that they missed the target speech (similar to the gated masker condition). Both gated and continuous maskers were tested because it is common in open-set sentence recognition to use a gated masker, but sentence-level semantic meaning of the masker speech may be more salient for a continuous masker.

Prior to experimental testing, listeners were familiarized with the task. For both groups, all listeners first heard sentences from list 20 of the BEL corpus. These sentences were presented in the presence of a two-talker masker that was randomly selected from among the four conditions (VS/M, SS/M, VS/A, or SS/A). The familiarization phase included (in this order): ten sentences presented at +5 dB signal-to-noise ratio (SNR), ten sentences presented at +2 dB SNR, and five sentences presented at −1 dB SNR. This process allowed the listener to become familiar with the task and the target talker's voice. Therefore, listeners had at least two strategies for identifying the target voice: they could use supra-threshold trials to learn the target talker's voice characteristics, and they could rely on temporal cues provided by the 500-ms delay between either masker onset (gated condition) or onset of the green light and the target (continuous condition).

Once the familiarization period was completed, experimental testing began. Experimental testing included four masker conditions (VS/M, SS/M, VS/A, and SS/A) and three SNRs (−1, −3, and −5 dB). One BEL list was presented for each masker condition at each SNR (25 sentences with 100 scorable keywords). The order of BEL lists was randomly selected for each listener, and listeners never heard the same list twice. Masker presentation order was randomly varied across participants. Within a masker condition, data were collected at increasingly negative SNRs (−1 dB SNR, followed by −3 dB SNR, followed by −5 dB SNR). A descending order was used to avoid floor effects in the most challenging SNR condition (e.g., Brouwer *et al.*, 2012; Calandruccio *et al.*, 2017a), as pilot testing indicated that performance for the −5 dB SNR condition was near floor performance for several listeners who had no prior listening experience at the more favorable SNRs. For all testing, the level of the target speech was fixed at 55 dB sound pressure level (SPL), resulting in an overall target + masker level of 58.5, 59.8, and 61.2 for the −1, −3, and −5 dB SNR conditions, respectively. Sentence recognition scores (number of keywords correct/100 possible keywords) were calculated for each condition (four masker conditions and three SNRs). These data were fitted with a logit function by minimizing the sum of squared error. A 50% sentence-recognition threshold (dB SNR) was calculated for each listener for each masker condition. Out of the 156 function fits, 6 were excluded because the function fit was poor (median $r^2 = 0.27$). This left 150 fits that were strong (median $r^2 = 0.97$).

D. Statistical methods

In order to examine the effects of both semantic content and syntactic structure of the masker speech, a regression analysis was conducted using a mixed linear model (*R* lme4; Bates *et al.*, 2015). The outcome of interest was sentence recognition threshold (dB SNR). The predictors were semantic meaning of masker sentences (meaningful or anomalous), syntactic structure of masker sentences (same or varied), and masker condition (gated or continuous). In our full model, we examined the three- and two-way interactions between semantic meaning, syntactic structure, and masker condition, as well as the associated main effects. Our data contain repeated measures within subjects. To account for correlations in the data due to these repetitions, random intercepts for each participant were included in the model. The level of statistical significance used in our analysis was $\alpha = 0.05$. *P*-values were adjusted for multiple testing when applicable and stated accordingly.

III. RESULTS

In our full regression model, the three-way interaction among semantic meaning, syntactic structure, and masker condition was not significant (*p*-value = 0.78). The two-way interaction between syntactic structure and masker condition was significant (*p*-value = 0.03), while the remaining two-way interactions were not significant (*p*-values = 0.26 and 0.60). We removed the three-way interaction term, as well as the two-way interaction terms, one at a time according to significance level and reassessed model fit. Variables were removed if they were not significant in the current model and if their removal did not significantly alter the amount of variability explained (Chi-squared test—see Table I for the final, reduced model). In our final, reduced model, a significant effect of semantic meaning (*p*-value < 0.001) and masker condition (*p*-value = 0.003) was found, as well as a significant two-way interaction between syntactic structure and masker condition (*p*-value = 0.007). The significant effect of semantic meaning was in the opposite direction to the prediction. On average, listeners had *lower* thresholds for the semantically meaningful (VS/M, SS/M) than the semantically anomalous masker conditions (VS/A, SS/A). This indicates that listeners' sentence recognition was better

TABLE I. Final-reduced mixed model logistic regression results for experiment 1.

	Estimate (β)	Standard Error	t ratio	df	P-value
Intercept[a]	−3.09	0.30	−10.3	48	<0.001
Semantic meaning (anomalous)	0.49	0.13	3.9	107	<0.001
Syntactic structure (same)	0.14	0.18	0.78	107	0.43
Masker condition (continuous)	−1.32	0.42	−3.14	44	0.003
Syntactic structure (same)× Masker condition (continuous)	−0.69	0.25	−2.74	107	0.007

[a] The intercept represents the baseline case of semantic meaning (meaningful), syntactic structure (varied), and masker condition (gated). Degrees of freedom are Satterthwaite approximations.

when listening to target speech in the presence of the semantically meaningful maskers than the semantically anomalous maskers with mean thresholds of −3.8 and −3.3 dB SNR, respectively (Fig. 3). The significant effect of masker condition reflects the fact that thresholds were higher in the gated masker than the continuous masker with mean thresholds of −2.7 and −4.4 dB SNR, respectively. This approximate 1.7 dB decrease in threshold for the continuous masker may be due to the listeners having access to the masker alone between trials, which could facilitate segregation of the target and masker into separate auditory streams.

Post hoc Tukey testing indicated that there was no effect of syntactic structure for the gated masker condition, but there was a significant effect of syntactic structure for the continuous masker condition: thresholds were higher for the *varied syntax* than the *syntax* conditions with mean thresholds of −4.2 and −4.7 dB SNR, respectively. This result is consistent with our hypothesis that a masker with a consistent rhythmic pattern is easier to segregate from the target and ignore than a masker with a varied rhythmic pattern. However, this interpretation should be treated with caution due to wide variability across participants and the small effect size of approximately 0.5 dB SNR.

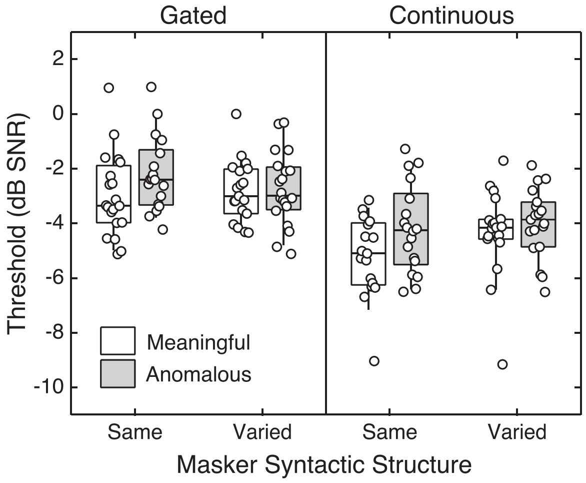

FIG. 3. Sentence-recognition thresholds from experiment 1 for meaningful target sentences in the presence of four different two-talker masker conditions (SS/M, SS/A, VS/M, VS/A). Data are shown for two groups of listeners. One group was presented stimuli using a gated masker paradigm, and the second group was presented stimuli using a continuous masker paradigm. Circles indicate individual listener thresholds. Box plots indicate the median value of the data (horizontal line within box) and the interquartile range (box length). Whiskers represent the 10th and 90th percentiles.

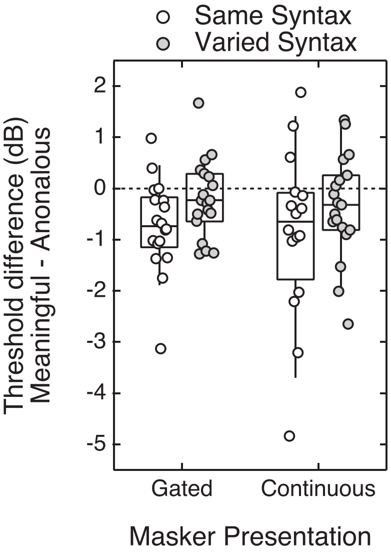

FIG. 4. Difference in sentence-recognition thresholds between meaningful and anomalous maskers from experiment 1. Differences were calculated by subtracting each individual listener's threshold in the semantically meaningful masker condition from the threshold for the semantically anomalous masker condition. Difference scores below the zero line indicate that listeners' thresholds were lower (better) in the meaningful masker condition relative to the semantically anomalous masker condition. Plotting conventions follow those of Fig. 3.

To highlight individual differences in thresholds between the semantically meaningful and anomalous masker conditions, Fig. 4 shows individual threshold differences between these masker conditions for all data included in the full model. The zero line on the y axis of Fig. 4 indicates no sentence-recognition threshold difference between the masker conditions that varied in semantic meaning. Although, on average, there is an advantage for meaningful maskers, the individual variability across listeners can be easily observed here with about a quarter of the participants showing no difference between maskers (meaningful vs anomalous) or even a slight advantage when listening in the presence of the anomalous masker condition (above the zero line).

IV. DISCUSSION

Experiment 1 evaluated whether masked-speech recognition varies depending on the sentence-level semantic content and predictability of the syntactic structure of the competing two-talker speech when controlling for lexical content and speaking rate between the maskers. The original hypotheses were that semantically meaningful maskers would be more effective than the semantically anomalous maskers, and maskers with varied syntax would be more effective than maskers with the same syntactic structure due to greater rhythmic uncertainty. Contrary to prior expectations, the *anomalous* (VS/A and SS/A) maskers proved to be more effective than the *meaningful* (VS/M and SS/M) maskers. Further, differences in syntactic structures only had an effect when the masker was played continuously throughout the experiment. When the masker was played continuously, thresholds were approximately 0.5 dB lower with the

SS maskers than the VS maskers. This result is generally consistent with the idea that the predictable rhythm associated with the SS condition may have benefitted performance. Further experimentation is needed to better understand this result.

The results from experiment 1 were not consistent with those observed in Brouwer *et al.* (2012). Since we did not observe the greatest masking when both the target and masker sentences were meaningful, these data do not support the linguistic similarity hypothesis with respect to the similarity within the semantic meaning of the target/masker speech. Further, the effect size associated with syntactic regularity may be too small to account for the masker effects observed by Brouwer *et al.* (2012). To better understand why our results did not align with those observed in Brouwer *et al.*, a second experiment was conducted using stimuli and methods that more closely resemble those used in that previous study.

A. Experiment 2: Similarity of semantic content between the target and masker speech

In Brouwer *et al.* (2012), the intention was not to assess the contributions of the semantic meaning of the masker speech, but rather to assess the importance of the linguistic similarity between the target and the masker with respect to the semantic meaningfulness. They explored performance differences for meaningful target sentences when presented in the presence of either meaningful masker sentences or semantically anomalous masker sentences. When there was a *mismatch* in semantic meaning (meaningful target + anomalous masker), performance improved. This result was explained as support for the *linguistic similarity hypothesis*. However, the complementary condition using semantically anomalous targets (anomalous target + meaningful masker) was not explored. For an anomalous target, the similarity hypothesis predicts more masking for a semantically anomalous masker than a meaningful masker.

It is well documented that the semantic meaningfulness of the *target* speech has a large effect on sentence recognition (e.g., Bradlow and Alexander, 2007; Kalikow *et al.*, 1977). The semantic context allows listeners to use higher level linguistic cues to fill in later words within a sentence based on the previously spoken words of the same sentence. It has also been shown that predictable syntax structures facilitate sentence recognition (Kidd *et al.*, 2014); when predictable syntax structures are maintained, sentence recognition is improved compared to when words are presented in a syntactical form not common to or not allowed in the target language.

The purpose of experiment 2 was twofold. First, to help us better understand our results from experiment 1 in relation to their data, we wanted to see if we could replicate the results observed in Brouwer *et al.* (2012) using the same speech materials used in that study, although recorded by different talkers. Second, we wanted to determine whether mismatching the target and masker speech with respect to semantic meaning would cause a change in masker effectiveness, testing both meaningful and anomalous targets, while also controlling for differences in syntactic structure within the masker speech.

V. METHODS

A. Participants

Participants for experiment 2 included 20 young-adult listeners (ages 18–31 yr, mean age = 20.7 yr, SD = 2.7 yr; 14 females, 6 males). Inclusion criteria with respect to normal hearing and native English-speaking status were the same as in experiment 1. Two participants had previously completed experiment 1.

B. Stimulus development

1. Target stimuli

All target stimuli were spoken by the same female talker, a white 29-yr-old female from Detroit, MI, with no noticeable regional accent. Target stimuli consisted of two types: semantically meaningful and semantically anomalous sentences. Semantically meaningful sentences were taken from the Bamford-Kowal-Bench (BKB) corpus (Bench *et al.*, 1979). These sentences have varied syntactic structures. Two examples from this sentence list are: "The CLOWN had a FUNNY FACE" and "STRAWBERRY JAM is SWEET." Sentences are scored based on keywords within the sentences, which are capitalized in the preceding examples. Thirteen different BKB lists (list 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 19, and 21) were used, following Brouwer *et al.* (2012). Each list contains 16 sentences (50 keywords/list). The average speaking rate for these recordings was 3.7 syllables/s; the average *f*0 was 205 Hz.

Semantically anomalous sentences were taken from the Helfer semantically anomalous sentences database (Helfer, 1997). These sentences also have varied syntactic structures. Two examples from this sentence database include "PUBLISH the FEATHER of the MORAL" and "The DREAM did REPLACE his THROAT." Each sentence has three keywords, as indicated by the capitalized words in the two previous examples. Sentences 1–250 were used. The average speaking rate for these recordings was 4.0 syllables/s, while the average *f*0 was 214 Hz. On average, the speaking rate of the anomalous sentences was faster than the meaningful targets [$t_{(1,418)} = 6.60$, $p < 0.001$], and the average fundamental frequency of the anomalous sentences was higher [$t_{(1,428)} = 11.96$, $p < 0.001$]. This corpus was chosen based on broad similarities to the BKB sentences with respect to lexical content, length, and syntactic complexity.

2. Masker stimuli

All maskers consisted of two-talker female speech. The same two female talkers were used to create all three masker conditions used in experiment 2. Talker A was a 19-yr-old Asian-American female from Dallas, TX, and talker B was a 20-yr-old white female from Cleveland, OH. Both masker talkers spoke standard American English, and neither talker had a detectable regional accent. The average *f*0 was 223 Hz and 208 Hz for talkers A and B, respectively.

The first two masker conditions used in this experiment were created to reassess the results observed in Brouwer *et al.* (2012); that is, we included one semantically meaningful masker, which was created using Harvard/Institute of Electrical and Electronics Engineers (IEEE) sentences (Egan, 1948; IEEE Subcommittee on Subjective Measurements, 1969), and one semantically anomalous masker, which was created using sentences from the Syntactically Normal Sentence Test (SNST; Nye and Gaitenby, 1974). The IEEE and SNST sentences chosen were identical to those used in Brouwer *et al.* (2012). An example of an IEEE sentence is "The blind man counted his old coins." An example of an SNST sentence is "The deep hurt cut the year." The IEEE corpus includes sentences with different syntactic structures, while the SNST sentences use a fixed syntactic structure. Based on this difference, and to follow a similar nomenclature as used in experiment 1, the IEEE masker will be referred to as the *varied syntax/meaningful* masker 2 (VS/M2), and the SNST will be referred to as the *same syntax/anomalous* masker 2 (SS/A2). The addition of the number "2" is to remind the reader that these are different maskers than those used in experiment 1. In addition to differences in semantic meaning and syntactic structure, there are other differences between the IEEE and SNST corpora. The IEEE sentences include monosyllabic and disyllabic keywords, while SNST sentences include only monosyllabic words. The IEEE corpus uses many unique words across maskers (732 unique/1545 total words in the subset of sentences used here), while the SNST uses significantly fewer unique words (253 unique/1194 total words). To evaluate some of the differences between the two maskers, a third masker condition was also included in experiment 2. The *same syntax/meaningful* masker 2 (SS/M2) consisted of semantically meaningful sentences with the same fixed syntactic structure as the SNST sentences and included only monosyllabic words. An example sentence from this masker was "The strong wind broke the fence."

For the VS/M2, talkers A and B read the same 100 IEEE sentences that were read by the talkers in Brouwer *et al.* (2012). All sentences were saved as individual WAV files and were edited to remove silent periods at the beginning or the end of the sentence. The average speaking rate for each individual masker stream was 3.6 and 3.7 syllables/s for talkers A and B, respectively.

For the *same syntax/anomalous* masker 2 (SS/A2), Talkers A and B read 100 sentences each from the SNST sentence materials. As for the semantically meaningful condition, these were the same exact sentences read by the masker talkers in the Brouwer *et al.* (2012) study. The semantically anomalous sentences have a fixed syntactic structure (DANVDN). The average speaking rate for each individual masker stream was 3.2 syllables/s and 3.8 syllables/s for talkers A and B, respectively.

The third masker condition, SS/M2, was created using 200 sentences from the BEL database used for target stimuli in experiment 1 (Calandruccio and Smiljanić, 2012). The sentences chosen from this database included semantically meaningful sentences that used the syntactic structure DANVDN, the same syntactic structure used for the SNST sentences. The 100 sentences from the BEL database were modified slightly to replace all multisyllabic words with monosyllabic words, and are referred to as modified-Basic English Lexicon sentences (MBEL) hereafter. The average speaking rate for this masker condition was 3.3 syllables/s and 3.7 syllables/s for talkers A and B, respectively. For the exact text used for the maskers created in experiment 2, please see supplemental material 3.[1]

The method for generating a two-talker masker based on sentence recordings for talkers A and B was consistent across conditions. Recordings were RMS equalized, and sentences spoken by each talker were concatenated to create two long wav files. The two WAV files, one for each masker talker, were then mixed together to create the two-talker masker stream.

C. Procedure

The procedure for experiment 2 was similar to that of experiment 1, however, similar to Brouwer *et al.* (2012), the masker was gated on and off on every trial with a 500 ms lead/lag time. In addition, in experiment 2, stimuli were presented diotically over headphones (Sennheiser HD-25, Wedemark, Germany). While insert earphones were used in experiment 1, circumaural phones were chosen for experiment 2 based on ease of use. Further, listeners were instructed that some sentences would be meaningful and some sentences would not make sense; in either case, the instructions were to determine the target voice and repeat back the entire target while trying to ignore the other two competing talkers. Listeners were told that the same three talkers would be used throughout the duration of the experiment, and the target voice would not change during testing. Throughout testing, the presentation level of the masker sentences was fixed at 60 dB SPL.

Testing was blocked by target condition (semantically meaningful and semantically anomalous), but the order of target condition was randomized across listeners. For each target condition, familiarization trials were completed prior to experimental testing. For the anomalous targets, familiarization trials consisted of 16 sentences from the set described above. Targets were presented at decreasing SNRs; sentences 1–5 played at +5 dB SNR, sentences 6–10 played at +3 dB SNR, and sentences 11–16 played at +1 dB SNR. For the meaningful targets, 16 BKB sentences from list 21 were presented at decreasing SNRs; sentences 1–5 played at +5 dB SNR, sentences 6–10 played at 0 dB SNR, and sentences 11–16 played at −1 dB SNR. The two-talker masker used in familiarization trials was randomly selected for each listener and each target condition.

After the familiarization trials, the experimental testing began. Meaningful target sentences were presented at −1 dB SNR followed by −3 dB SNR. The semantically anomalous target sentences were presented at +1 dB SNR followed by −1 dB SNR. Testing within a target condition was blocked by masker type. The differences in SNR between the semantically meaningful and semantically anomalous target sentences were based on pilot testing and reflect the increased difficulty listeners had recognizing the anomalous sentences.

Target sentence lists were randomly chosen for each condition and listener, with the proviso that listeners never heard the same target sentence twice. Sentence recognition scores (number of keywords correct/approximately 100 possible keywords) were calculated for each condition. The specific number of possible keywords varied depending on the meaningfulness of the targets; two lists of meaningful sentences were used per condition (approximately 96 keywords), and 34 anomalous sentences were used per condition (approximately 102 keywords). These data were fitted with a logit function by minimizing the sum of squared error. A 50% sentence-recognition threshold (dB SNR) was calculated for each listener for each target and masker condition (a total of 120 fits). Data from one participant were excluded due to low reliability and extremely poor performance across conditions. In the remaining data, six data points were omitted due to poor function fits.[2] This left 114 estimates of SRT.

D. Statistical methods

In order to examine the effects of semantic content of the target and masker conditions, a regression analysis was conducted using a linear-mixed model (*R* lme4). The statistical approach employed in experiment 1 was followed here. The outcome of interest was sentence recognition threshold (dB SNR). The predictors were target type, which was either meaningful or anomalous, and masker condition, which was either VS/M2 (IEEE), SS/M2 (MBEL), or SS/A2 (SNST). In our full model, we examined the two-way interaction between target type and masker condition and the associated main effects. The data contained repeated measures, and random intercepts for each participant were included in the model. The level of statistical significance used in our analysis was $\alpha = 0.05$. *P*-values were adjusted for multiple testing when applicable and stated accordingly.

VI. RESULTS

As can be seen in Fig. 5, there were minimal differences across masker conditions within each target condition. We also can observe that performance was poorer overall for anomalous targets. These observations were confirmed by the statistical analysis, which indicated no significant main effect of masker condition ($p = 0.84$), nor a significant interaction ($p = 0.40$), but a significant main effect of target type ($p < 0.001$) (see Table II for full model results). The average threshold was -1.9 dB for meaningful and 0.5 dB for anomalous sentences for an average difference between the two target types of 2.4 dB. Even though we used the same stimulus materials as used by Brouwer *et al.* (2012; meaningful BKB target sentences paired with either the VS/M2 IEEE masker sentences or the SS/A2 SNST masker sentences), we failed to replicate their results. However, a notable difference between the statistical analysis employed by Brouwer *et al.* (2012) and the one used here was that the previous study included each stimulus item as a random effect in the statistical model. An additional analysis was therefore conducted.

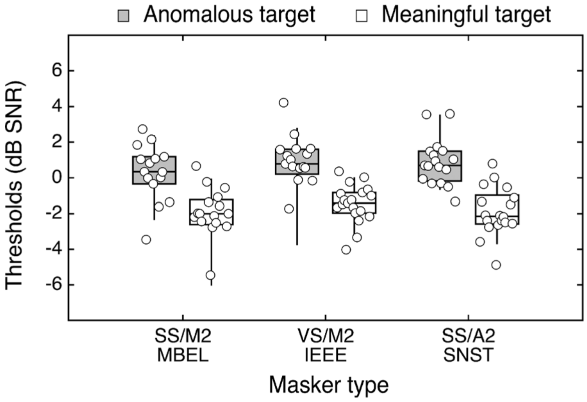

FIG. 5. Sentence-recognition thresholds for anomalous and meaningful speech targets for three different two-talker masker conditions from experiment 2 (SS/M2, VS/M2, SS/A2). Plotting conventions follow those of Fig. 3.

TABLE II. Full mixed model logistic regression results for experiment 2 (sentence recognition thresholds).

	Estimate (β)	Standard Error	*t* ratio	*P*-value
Intercept[a]	0.77	0.46	1.67	0.10
Target (meaningful)	-2.66	0.53	-4.97	<0.001
Masker [VS/M2 (IEEE)]	-0.44	0.54	-0.82	0.41
Masker [SS/M2 (MBEL)]	-0.34	0.56	-0.60	0.44
Target (meaningful) × Masker(VS/M2)	0.88	0.75	1.17	0.24
Target (meaningful) × Masker (SS/M2)	0.02	0.77	0.02	0.98

[a] The intercept represents the reference case target (anomalous) and masker [SS/A2 (syntactically normal sentence test, SNST)].

A regression analysis was conducted using a generalized linear mixed model (*R*, lme4) with a logistic link function. In this additional analysis, the dependent variable was word recognition for each word within the target sentence (correct/incorrect). Due to the binary nature of the outcome variable, odds ratios were calculated. Odds ratios characterize the chance of getting a word correct. Determining odds ratios allows us to compare odds between two conditions. The predictor variables included in the generalized linear mixed model were SNR (easy or difficult), meaningfulness of target sentence (meaningful or anomalous), and masker sentence type [VS/M2 (IEEE), SS/A2 (SNST), or SS/M2 (MBEL)]. We removed variables from our full model that were not significant and did not explain any additional variability in our outcome (*F*-test). Our data contain repeated measures within subjects, with multiple sentences and multiple words per sentence repeated across subjects. To account for correlations in the data due to these repetitions, we adjusted for random intercepts due to subject, sentence, and words nested within sentences as random effects in our model. The inclusion of words nested within sentences did not significantly alter the results of the regression analysis.

The three-way interaction involving SNR, target type, and masker type was not significant (*p*-value $= 0.95$), and was removed from the model. Once removed, the subsequent model indicated a lack of significant two-way interactions, and therefore the final model did not include these interactions (*p*-values ranging from 0.23 to 0.63). However, the main effects of SNR (easy vs difficult), target type (meaningful vs anomalous), and masker (VS/M2 vs SS/A2 vs SS/M2) were significant (*p*-value < 0.001) and retained in our final model.

TABLE III. Final-reduced mixed model logistic regression for experiment 2 (word identification results).

	Odds ratio (β)	Standard Error	z-statistic	P-value
Intercept[a]	1.10	0.16	0.55	0.59
SNR (difficult)	0.32	0.03	−37.38	<0.001
Target [meaningful (BKB)]	1.33	0.07	4.28	<0.001
Masker (VS/M2)	0.76	0.04	−7.33	<0.001
Masker (SS/M2)	0.997	0.04	−0.09	0.93

[a]The intercept represents the reference case SNR (easy), masker [SS/A2 (SNST)], and target (anomalous).

The reference case in our final, reduced model was semantically anomalous target sentences, the same syntax/semantically anomalous masker [SS/A2 (SNST)], and the easy SNR condition (see Table III). The odds of correctly identifying a word in the reference case was 1.10 (10% more likely to be correct than incorrect). The effect of switching from an easy SNR to a difficult SNR (a 2 dB change) was to decrease odds by $\sim$68%. Odds increased by $\sim$33% if the target sentences were semantically meaningful vs semantically anomalous. In order to examine the effect of masker condition in the absence of interactions, we performed pairwise contrasts for masker condition averaged over target condition and SNR.

Table IV shows the odds ratios for each of the masker combinations. For the semantically meaningful maskers, the odds ratio was 31% higher for the syntactically same (SS/M2; MBEL) than for the syntactically varied (VS/M2; IEEE) masker sentences. That is, the VS/M2 sentences provided significantly more masking than the SS/M2 masker, even though both maskers consisted of semantically meaningful sentences. For the syntactically same maskers, the odds ratio was not significantly different for semantically meaningful (SS/M2; MBEL) and anomalous (SS/A2; SNST) masker sentences. Finally, using semantically meaningful, but syntactically varied (VS/M2; IEEE) sentences decreased odds by 23% compared to sentences that were semantically anomalous but syntactically the same (SS/A2; SNST).

VII. DISCUSSION

When sentence-recognition threshold was the outcome variable of interest, no significant differences were observed across masker conditions, nor was there an interaction between target and masker meaning. The lack of an interaction does not support the idea that sentence-level semantic meaning is relevant with respect to the *linguistic similarity hypothesis*. However, when the power of the statistical model was increased with the analysis of responses by word within each sentence, the VS/M2 masker was significantly more effective than the SS/A2 masker. This result is in line with that reported by Brouwer *et al.* (2012; VS/M2 > SS/A2). However, the data also suggest that semantic meaning of the varied syntax masker (VS/M2) is not likely the cause of its effectiveness. When masking was compared for semantically meaningful and anomalous sentences with similar (fixed) syntactic structure [SS/M2 (MBEL) and SS/A2 (SNST)], no difference in masker effectiveness was observed. Further, the VS/M2 masker was more effective than the SS/M2 masker, despite the fact that both maskers were semantically meaningful.

TABLE IV. Pairwise least squares means contrasts for masking condition in the final, reduced model. (Note: P-values and confidence intervals have been adjusted for multiple testing using the Tukey method.)

Contrast	Odds ratio	95% CI	Standard Error	Z-statistic	P-value
SS/M2 - VS/M2	1.31	[1.20,1.43]	0.05	7.33	<0.001
SS/M2 - SS/A2	1.00	[0.92,1.09]	0.04	0.09	0.996
VS/M2 - SS/A2	0.77	[0.70,0.84]	0.03	−7.20	<0.001

Overall, our results do not support the hypothesis that greater similarity between the target and masker speech, with respect to sentence-level semantic meaning, results in more masking. Therefore, contrary to conclusions of Brouwer *et al.* (2012), it appears that the linguistic similarity hypothesis does not hold with respect to sentence-level semantic meaning between the target and masker speech.

The finding of a significant effect of masker using an item-level analysis highlights the importance of making a distinction between a statistically significant effect and a meaningful difference. For the sentence-level model with a relatively small number of data points ($\sim$30 sentences per listener and condition), no difference in masker condition was observed. However, for the word-level model with a larger number of data points (up to $\sim$200 words per listener and condition), a statistical difference was observed, indicating that the VS/M2 (IEEE) masker was most effective. This significant effect was apparent even after removing item as a random factor in our statistical model, indicating it was the increase in the number of data points, not controlling for item (each specific word) in the model, that provided the increased power for the second (binary word identification) analysis. The overall mean threshold difference among the three masker conditions ranged between 0.1 (VS/M2 – SS/A2) and 0.3 (VS/M2 – SS/M2) dB. These threshold differences are very small, and likely not meaningful differences among the masker conditions. Median values of threshold and slope from the data fits were used to estimate the effect size in terms of the change in percent correct. At −2 dB SNR, BKB sentence data were consistent with performance differences ranging from 7.4 percentage points (VS/M2 – SS/A2) to 9.7 percentage points (VS/M2 – SS/M2). A close examination of the data reported by Brouwer *et al.* (2012) also indicates rather small effects of masker type. The mean performance difference between the meaningful and anomalous maskers in that study was 8 percentage points [see Brouwer *et al.*, 2012, p. 1455, Fig. 2(B)].

A. General discussion

The purpose of these experiments was to re-examine whether semantic meaning of the masker speech influences masked-speech recognition while taking syntactic structure of the masker speech into consideration. In experiment 1, which utilized meaningful target sentences, semantic content and syntactic structure of the masker speech was varied (semantically meaningful vs semantically anomalous and varied syntax structures vs one fixed syntax structure), while controlling for lexical content and speaking rate. One hypothesis at the outset was that a semantically meaningful masker would be more effective than a semantically anomalous masker. Contrary to the linguistic similarity hypothesis, for semantically meaningful target sentences, more masking was observed for the semantically anomalous masker than the meaningful one, although the effect was small (0.6 dB). It is hard to make sense of this result. Experiment 1 is the first study of semantic meaning of the masker speech to employ strict control over other linguistic features (syntax, speaking rate, lexicon, syllable count). Though more research is needed to understand this effect, it might suggest that despite our efforts to control the acoustical parameters between the speech maskers, acoustical differences between the masker streams still remain that change listeners' ability to segregate the target from the masker speech.

A second hypothesis evaluated in experiment 1 was that masker speech composed of sentences with identical syntax structures would facilitate sentence recognition by providing listeners with a predictable rhythmic pattern and limiting uncertainty, thereby facilitating segregation of the target and masker. A significant interaction was observed between syntax (same and varied) and masker condition (gated and continuous). Specifically, when listeners heard masker speech continuously, rather than gated, having the SS across all masker sentences facilitated sentence-recognition and decreased thresholds by ~0.5 dB. Further evidence is needed to draw any strong conclusions from this finding; however, it is possible that the continuous masker allowed listeners more time to form an auditory stream and utilize the predictable rhythm of the masker to aid in segregation of the target and masker speech (Moore and Gockel, 2012). While Brouwer *et al.* (2012) used a gated masker, the finding of more masking with irregular syntactic structure of the masker could bear on the interpretation of their results; that is, more masking with the semantically meaningful and syntactically irregular masker could be due to syntactic rather than semantic features, a possibility that undermines the linguistic similarity hypothesis with respect to sentence-level semantic meaning.

Experiment 2 was conducted in an attempt to better understand the results of experiment 1 and discrepancies with the results of Brouwer *et al.* (2012). Thus, we used the same speech materials as that previous study, although our recordings used different talkers. Some data indicate that the combination of different talkers' voices and what they are saying can greatly influence the amount of masking (Boulenger *et al.*, 2010; Calandruccio *et al.*, 2017b; Helfer and Jesse, 2015), while other studies show minimal differences between masker conditions (Brungart *et al.*, 2005; Jones and Litovsky, 2008). Further, the effect that Brouwer *et al.* (2012) observed was small (similar to our results of experiment 1). As in experiment 1, when syntactic structure and syllable count were controlled, differences in semantic meaning of the masker did not impact masked-speech recognition (SS/M2 vs SS/A2). When sentence-recognition thresholds were examined, no differences in any masker conditions were found. However, when the power of the statistical model was increased by analyzing performance by word, we were able to observe a significant effect of masker: the meaningful masker with varied syntax (VS/M2) was the most effective masker of the three. This result is also broadly consistent with the idea that the results of Brouwer *et al.* could reflect effects of syntactic features of the masker rather than similarity of the target and masker with respect to semantic meaning.

B. Meaningfulness of masker speech does not impact masked-speech recognition

The most notable finding of the current data set is that the meaningfulness of the masker speech does not increase masking. In both experiments 1 and 2, there was no increase in masking due to meaningful sentence-level semantic content. In experiment 1, greater masking was actually observed for maskers that were created using sentences that were semantically anomalous. In experiment 2, increasing the power of the statistical model by including 1200 data points/listener indicated significantly more masking for the varied syntax meaningful masker (VS/M2) compared to the anomalous masker with the same, fixed syntax structure (SS/A2). However, this difference did not generalize to the other semantically meaningful masker (SS/M2), casting doubt that the semantic meaning was the reason for its greater effectiveness.

The current data set adds to the growing number of examples in the literature that support the idea that the meaningfulness of the masker speech is not critical with respect to masking. For example, Calandruccio *et al.* (2010a) reported that English masked-speech recognition performance did not significantly vary when listening in two different non-native accented two-talker maskers competing in the background. That is, performance was not statistically different despite significant differences in the intelligibility of the competing-accented speech, which ranged from approximately 44% to 88% correct. In a different example, simultaneous bilinguals with high levels of language proficiency in both English and Greek, showed large improvements in English masked-speech recognition with competing three-talker Greek speech, compared to an English masker. The linguistic masking release observed for the simultaneous bilinguals was similar in magnitude for monolingual speakers of English with no knowledge of the Greek language (Calandruccio and Zhou, 2014). A similar result was observed for bi-dialectical speakers of Dutch and Limburgian (Brouwer, 2017). Though the Dutch-Limburgian native speakers had high levels of listening and speaking abilities in both Dutch and Limburgian dialects, when Limburgian masker speech was competing in the background, Dutch native and Dutch-Limburgian native listeners benefited to the same degree when listening to Dutch target sentences compared to when Dutch speech was the masker. These data support the idea that the mismatch in language between the target and the masker, not the meaningfulness of the masker language, is what listeners use to segregate the target from the masker and improve their overall speech recognition.

The finding that the meaningfulness of the masker speech is not important with respect to masking is also supported by data from adults and school-aged children showing no significant difference in masked-speech recognition when the masker was meaningful speech or jumbled speech (Newman *et al.*, 2015). In this experiment, words were individually recorded and then concatenated to create masker speech. The resulting word sequences were either meaningful (e.g., "Airplanes. Fly. High. And. Quite. Fast.") or anomalous (e.g., "High. Fast. Fly. Airplanes. And. Quite."). The authors cautioned that the lack of coarticulation and sentence-level prosody could have impacted their data. However, their results are consistent with data from experiment 1 of the present report, which utilized natural speech recordings of coarticulated semantically anomalous speech. Newman *et al.* (2015) concluded that masker speech is likely to cause lexical competition/confusion, but that the listener is not processing the sentence-level semantic content of the competing speech.

C. Implications for the linguistic similarity hypothesis

In Brouwer *et al.* (2012, p. 1449) it was stated that, "The target-masker linguistic similarity hypothesis assumes that the more similar the target and the masker speech, the harder it is to segregate the two streams effectively." This general principle is supported by data on target/masker differences with respect to sex (Festen and Plomp, 1990; Helfer and Freyman, 2008; Leibold *et al.*, 2018), time reversal of the masker speech (Calandruccio *et al.*, 2017a; Iyer *et al.*, 2010; Rhebergen *et al.*, 2005), language (Bradlow and Alexander, 2007; Calandruccio *et al.*, 2010b; García Lecumberri and Cooke, 2006; Van Engen, 2010), accent (Calandruccio *et al.*, 2010a), phonetic distance between languages (Calandruccio *et al.*, 2013), and dialect (Brouwer, 2017). In all of the examples that support this hypothesis, however, there are significant differences in the acoustic or phonetic features between the target-masker pairings. For example, time reversing a speech masker introduces sound sequences that do not appear in natural speech (e.g., a reverse plosive). It has been suggested by Hoen *et al.* (2007) that informational masking for speech-on-speech consists of two components: acoustic-phonetic masking and lexical masking. Boulenger *et al.* (2010) reported significantly greater masking for speech babble composed of words with high-lexical frequency rather than low-lexical frequency. However, they also reported significant differences between the acoustic/phonetic features of the two masker streams. Our data, taken together with other literature referenced above, imply that sentence-level semantic content is less important (or not important at all) once significant acoustic-phonetic differences between the target-masker speech are removed or minimized.

Lexical competition at the word level could explain why even simultaneous bilinguals who are fluent in both target and masker speech languages benefit to the same degree as monolinguals when listening in a target/masker mismatched language paradigm. It should be noted that it has been observed that non-native speakers of English experience a *smaller* masking release compared to monolinguals when listening to English in the presence of their native language (Brouwer *et al.*, 2012; Van Engen, 2010). However, the smaller masking release experienced by non-native speakers may have to do with the fact that the sequential bilinguals in those studies had relatively poor masked-speech recognition in their second language in general. Non-native speakers may require access to more low-level speech cues to accurately recognize speech than native speakers (Calandruccio and Buss, 2017) such that they continue to struggle even when the target and masker are segregated. This would tend to minimize their potential for linguistic-masking release. However, a comprehensive model of non-native speech perception is yet to be developed.

The data of Dai *et al.* (2017) are also relevant to the existence of lexical masking. That study tested masked-speech recognition with a vocoded speech masker. At first the masker speech was unintelligible to the participants, but after training, listeners were able to recognize the vocoded speech with upward of 56% correct. This increase in masker intelligibility was associated with a decrease in listeners' ability to recognize target speech presented with that vocoded masker. This result was argued to reflect increased lexical competition with increased masker intelligibility. This supports the idea that lexical competition between the target and the masker hurts performance. An alternative explanation for the results of Dai *et al.* (2017) is that training with the vocoded speech masker improved listeners' ability to recognize phonemes within the vocoded speech. It is therefore possible that phonetic confusions decreased the overall intelligibility rather than lexical interference.

In our data set, sentence-level semantic meaning was not an important factor in masker effectiveness. Due to the acoustic/phonetic similarities between the target and masker speech streams, our masked-sentence recognition task was difficult, but the difficulty did not increase based on the meaningfulness of the masker speech. Further, the current data indicate lower thresholds when masker speech has a fixed pattern of syntax structures than varied syntax, a result interpreted as reflecting reduced stimulus uncertainty due to the predictable rhythmic syllabic structure. It should be noted, however, that this effect was small and may differ depending on the specific stimuli used (e.g., one- vs two-talker maskers) and the population of listeners tested (e.g., children vs older adults, musicians vs non-musicians). That is, differences in semantic meaning or rhythm between the maskers might be more noticeable to different listener populations and in turn have a larger effect. More research is needed to determine if word-level semantic content is critical, or whether phonemic interference is the driving force of a listener's perceptual difficulty in competing talker situations.

ACKNOWLEDGMENTS

Funding support was provided by the National Institutes of Health Grant No. R03 DC015074. We are thankful to Susan Mason and acknowledge her efforts during the initial work on this project.

[1] See supplementary material at https://doi.org/10.1121/1.5081679 for the sentences used to create the VS/M masker condition (supplementary material 1), all of the text for the masker conditions used in experiment 1 (supplementary material 2), and all of the text for the masker conditions used in experiment 2 (supplementary material 3).

[2] Because there were only two data points/conditions, we did not use an r^2 criterion. Fits in which performance did not decrease with decreases in SNR (as expected) were not included.

American Speech-Language-Hearing Association. (**2005**). "Guidelines for manual pure-tone threshold audiometry," available at www.asha.org/policy (Last viewed February 1, 2016).

Baayen, R. H., Piepenbrock, R., and Van Rijnm, H. (**1995**). "The CELEX lexical database. Release 2 (CD-ROM)" (Linguistic Data Consortium, University of Pennsylvania, Philadelphia).

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Bench, J., Kowal, Å., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Brit. J. Audiol. **13**(3), 108–112.

Billig, A., Davis, M. H., Deeks, J. M., Monstrey, J., and Carlyon, R. P. (**2013**). "Lexical influences on auditory streaming," Curr. Biol. **23**(16), 1585–1589.

Boulenger, V., Hoen, M., Ferragne, E., Pellegrino, F., and Meunier, F. (**2010**). "Real-time lexical competitions during speech-in-speech comprehension," Speech Commun. **52**(3), 246–253.

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**(4), 2339–2349.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA).

Brouwer, S. (**2017**). "Masking release effects of a standard and a regional linguistic variety," J. Acoust. Soc. Am. **142**(2), EL237–EL243.

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., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**(1), 292–304.

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., and Buss, E. (**2017**). "Spectral integration of English speech for non-native English speakers," J. Acoust. Soc. Am. **142**(3), 1646–1654.

Calandruccio, L., Buss, E., and Bowdrie, K. (**2017a**). "Effectiveness of two-talker maskers that differ in talker congruity and perceptual similarity to the target speech," Trends Hear. **21**, 1–14.

Calandruccio, L., Buss, E., Leibold, L., and Lowery, M. (**2017b**). "Stimulus features affecting speech recognition in a two-talker masker," J. Acoust. Soc. Am. **141**(5), 3819–3819.

Calandruccio, L., Dhar, S., and Bradlow, A. R. (**2010a**). "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., Gomez, B., Buss, E., and Leibold, L. J. (**2014**). "Development and preliminary evaluation of pediatric Spanish-English speech perception task," Am. J. Audiol. **23**(2), 158–172.

Calandruccio, L., and Smiljanić, R. (**2012**). "New sentence recognition materials developed using a basic non-native English lexicon," J. Speech Lang. Hear. Res. **55**(5), 1342–1355.

Calandruccio, L., Van Engen, K. J., Dhar, S., and Bradlow, A. R. (**2010b**). "The effectiveness of clear speech as masker," J. Speech Lang. Hear. Res. **53**(6), 1458–1471.

Calandruccio, L., and Zhou, H. (**2014**). "Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance," J. Speech Lang. Hear. Res. **57**(3), 1089–1097.

Carhart, R., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryngol. **91**(3), 273–279.

Carlile, S. (**2014**). "Active listening: Speech intelligibility in noisy environments," Acoust. Austral. **42**(2), 90–96.

Culling, J. F., and Stone, M. A. (**2017**). "Energetic masking and masking release," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, Springer Handbook of Auditory Research 60 (Springer International Publishing, Switzerland), pp. 41–73.

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.

Davis, C. J. (**2005**). "N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics," Behav. Res Methods **37**(1), 65–70.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**(1), 368–379.

Egan, J. P. (**1948**). "Articulation testing methods," Laryngoscope **58**(9), 955–991.

Elliot, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (**1979**). "Children's understanding of monosyllabic nouns in quiet and in noise," J. Acoust. Soc. Am. **66**(1), 12–21.

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., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**(2), 1040–1046.

García Lecumberri, M. L., 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.

Gordon-Salant, S., and Cole, S. S. (**2016**). "Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing," Ear Hear. **37**(5), 593–602.

Hall, J. W., Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-taker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**(2), 159–165.

Helfer, K. S. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech," J. Speech Lang. Hear. Res. **40**(2), 432–443.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**(1), 87–98.

Helfer, K. S., and Jesse, A. (**2015**). "Lexical influences on competing speech perception in younger, middle-aged, and older adults," J. Acoust. Soc. Am. **138**(1), 363–376.

Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Lang. Hear. Res. **33**(1), 149–155.

Hillock-Dunn, A., Taylor, C., Buss, E., and Leibold, L. J. (**2015**). "Assessing speech perception in children with hearing loss: What conventional clinical tools may miss," Ear Hear. **36**(2), e57–e60.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., and Collet, L. (**2007**). "Phonetic and lexical interferences in informational masking during speech-in-speech comprehension," Speech Commun. **49**(12), 905–916.

IEEE Subcommittee on Subjective Measurements (**1969**). "IEEE recommended practice for speech quality measurements. Standards Publication No. 297," IEEE Trans. Audio Electroacoustics **17**(3), 225–246.

Iyer, N., Brungart, D. S., and Simpson, B. D. (**2010**). "Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task," J. Acoust. Soc. Am. **128**(5), 2998–3010.

Jakien, K. M., Kampel, S. D., Stansell, M. M., and Gallun, F. J. (**2017**). "Validating a rapid, automated test of spatial release from masking," Am. J. Audiol. **26**(4), 507–518.

Jerger, J., Speaks, C., and Tramell, J. L. (**1968**). "A new approach to speech audiometry," J. Speech Hear. Disord. **33**(4), 318–328.

Jones, G. L., and Litovsky, R. Y. (**2008**). "Role of masker predictability in the cocktail party," J. Acoust. Soc. Am. **124**(6), 3818–3830.

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**(5), 1337–1361.

Kidd, G., Jr., and Colburn, H. S. (**2017**). "Informational masking in speech recognition," in *The Auditory System at the Cocktail Party*, Springer Handbook of Auditory Research 60, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer International Publishing, Switzerland), pp. 75–109.

Kidd, G., Mason, C. R., and Best, V. (**2014**). "The role of syntax in maintaining the integrity of streams of speech," J. Acoust. Soc. Am. **135**(2), 766–777.

Leibold, L. J., Calandruccio, L., and Buss, E. (**2018**). "Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch," Ear Hear. **39**(5), 935–945.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second language acquisition of and perception of speech in noise," J. Speech Lang. Hear. Res. **40**(3), 686–693.

McDermott, J. H., Wrobleski, D., and Oxenham, A. J. (**2011**). "Recovering sound sources from embedded repetition," Proc. Natl. Acad. Sci. U.S.A. **108**(3), 452–463.

Moore, B. C. J., and Gockel, H. E. (**2012**). "Properties of auditory stream formation," Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**, 919–931.

Newman, R. S., Morini, G., Ahsan, F., and Kidd, G. (**2015**). "Linguistically based informational masking in preschool children," J. Acoust. Soc. Am. **138**(1), EL93–EL98.

Nye, P. W., and Gaitenby, J. H. (**1974**). "The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences," Haskins Laboratories Status Report on Speech Research, SR-37/38, pp. 169–190.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Lang. Hear. Res. **29**(2), 146–154.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**(3), 1274–1277.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**(1), 112–117.

Van Engen, K. J. (**2010**). "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble," Speech Commun. **52**(11-12), 943–953.