Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content

Susanne Brouwer[a)]
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

Kristin J. Van Engen
Department of Linguistics, University of Texas, 1 University Station B5100, Austin, Texas 78712

Lauren Calandruccio
Department of Linguistics and Communication Disorders, Queens College of the City University of New York, 65-30 Kissena Boulevard, Flushing, New York 11367

Ann R. Bradlow
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 9 June 2011; revised 5 October 2011; accepted 13 December 2011)

This study examined whether speech-on-speech masking is sensitive to variation in the degree of similarity between the target and the masker speech. Three experiments investigated whether speech-in-speech recognition varies across different background speech languages (English vs Dutch) for both English and Dutch targets, as well as across variation in the semantic content of the background speech (meaningful vs semantically anomalous sentences), and across variation in listener status vis-à-vis the target and masker languages (native, non-native, or unfamiliar). The results showed that the more similar the target speech is to the masker speech (e.g., same vs different language, same vs different levels of semantic content), the greater the interference on speech recognition accuracy. Moreover, the listener's knowledge of the target and the background language modulate the size of the release from masking. These factors had an especially strong effect on masking effectiveness in highly unfavorable listening conditions. Overall this research provided evidence that that the degree of target-masker similarity plays a significant role in speech-in-speech recognition. The results also give insight into how listeners assign their resources differently depending on whether they are listening to their first or second language.
 [DOI: 10.1121/1.3675943]

PACS number(s): 43.71.Es, 43.71.Hw, 43.72.Dv, 43.71.Sy [MAH] Pages: 1449–1464

I. INTRODUCTION

In daily life, a challenge for interlocutors is to understand each other in spite of the presence of a variety of background noises. These competing noises may or may not contain linguistic information themselves. Pollack (1975) proposed a distinction between energetic and informational masking (see also Carhart *et al*., 1969, and Kidd *et al*., 2007, for a review). Energetic masking refers to masking at the auditory periphery and is related to the audibility of the target signal. This type of masking produces partial loss of information due to spectral and temporal overlap between the noise and the signal. Informational masking refers to the masking beyond what can be attributed to energetic masking alone. In the case of informational masking, the target and the noise may both be audible, but they may be difficult to separate, thus interfering with the recognition of the target. Informational masking therefore depends on factors that inhibit or facilitate stream segregation including linguistic, attentional, and other cognitive factors (Mattys *et al*., 2009). The present study explores the linguistic component of informational masking during speech-in-speech recognition.

We hypothesize that speech-on-speech masking is sensitive to variation in the degree of linguistic similarity between the target and masker speech, particularly under relatively unfavorable signal-to-noise ratios (SNRs). 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. Conversely, the hypothesis assumes that the more dissimilar the target and the masker speech, the easier it is to segregate the two streams effectively. The intuition behind this hypothesis can be illustrated on the basis of language-related, stimulus-related, and listener-related factors. For example, linguists and non-linguists are all likely to agree that Dutch and German are each more similar to English than is either Mandarin or Korean, yet the degree of similarity of Mandarin and Korean to English are harder to assess (see Bradlow *et al*., 2010, for further discussion of this issue). Nevertheless, we presume that target-masker language pairs can indeed be ranked relative to each other in terms of target-to-masker similarity.

We also presume that the relative similarity between two speech streams will depend on stimulus-related factors, such as the phonetic, semantic, and/or syntactic content of

[a)]Author to whom correspondence should be addressed. Present address: Northwestern University, Department of Linguistics, 2016 Sheridan Road, Evanston, IL 60208-4090. Electronic mail: s-brouwer@northwestern.edu

the target and masker stimuli. For example, linguistic interference will be more likely to appear in conditions in which the target and the masker tap into the same level of processing (e.g., semantically meaningful targets in semantically meaningful maskers) than into different levels of processing (e.g., semantically meaningful targets in semantically anomalous maskers).

Finally, the relative similarity between target and masker will depend on listener-related factors such as listener knowledge/experience with the target and masker language(s). For example, intelligible maskers will be more detrimental to target recognition than unintelligible maskers (cf., Van Engen and Bradlow, 2007). Thus, in general, the target-masker linguistic similarity hypothesis claims that a significant predictor of speech-in-speech recognition accuracy is target-masker similarity along these linguistically defined dimensions.

A competing hypothesis is that the above mentioned factors, especially the language- and stimulus-related factors, play a secondary (if any) role in accounting for variation in speech-on-speech masking. Under this view, observed masking differences across various speech maskers are attributable to informational masking differences that are not specific to linguistic processes or representations (i.e., attentional or cognitive components that do not depend on variation in phonetic, syntactic or semantic content) or to general energetic factors (i.e. spectral and/or temporal target-masker overlap). For example, this hypothesis presumes that speech recognition differences in same-language (e.g., English-in-English) versus different-language (e.g., English-in-Mandarin) target-masker combinations reflect general energetic and/or attentional-cognitive masking differences.

Teasing apart these two alternative hypotheses about the role of linguistic factors in speech-in-speech recognition is difficult because factors that decrease similarity between a target and masker likely also increase their general (i.e., non-linguistic) acoustic difference. Nevertheless, we seek to establish an independent contribution to speech-on-speech masking of variation along a linguistically defined distance dimension, that is, along a higher-order dimension of speech signal differentiation that involves a combination of phonetic, semantic and listener-related factors. This goal is potentially highly relevant for understanding native- and second-language speech perception development where the relative time-courses of general auditory and linguistic stream segregation may be misaligned. That is, stream segregation based on target-masker linguistic similarity and stream segregation based on general auditory distance may be two separate skills that may differ in the rate and/or manner in which they develop, decline and respond to training. Thus, our aim is to test speech-in-speech recognition under a variety of target-masker linguistic similarity conditions. We attempt to attenuate the effects of the inevitable concomitant variations in general (i.e., non-linguistic) auditory distance by equating the long-term average speech spectrum (LTAS) of the maskers (a speech signal manipulation that has a negligible effect on its intelligibility) and by comparing relative masking effectiveness across SNRs. The LTAS normalization helps reduce the effect of differences in purely energetic masking across speech maskers, and comparison across SNRs may highlight differences across various maskers under conditions of constant energetic masking differences.

In the present study, we investigated whether speech-in-speech recognition varies across (1) different background speech languages (English vs Dutch) for both English and Dutch targets (i.e., the signal of interest), (2) across variation in the semantic content of the background speech (meaningful vs semantically anomalous sentences), and (3) across variation in listener status vis à vis the target and masker languages (native, non-native, or unfamiliar). In a companion study (Calandruccio *et al.*, 2009), we focused on target-masker typological distance by including three masker languages (English, Dutch, and Mandarin) and three listener groups (English, Dutch, and Mandarin) that vary in their linguistic typological distances from the target language (English). Together, these studies allow us to observe speech-in-speech recognition under conditions of varying target-masker linguistic distance as defined with respect to the languages, the linguistic content, and the listener's native or non-native status.

A number of studies have looked at effects of background language on the recognition of sentences in the listener's native language. These studies found that performance decreases when the competing speech is spoken in the listeners' native language versus a language that is foreign or unfamiliar to them (e.g., Calandruccio *et al.*, 2010; Garcia-Lecumberri and Cooke, 2006; Rhebergen *et al.*, 2005; Van Engen and Bradlow, 2007). For instance, native English listeners received a release from masking (i.e., reduction in masking) when recognizing English sentences in the presence of two-talker Mandarin versus English background speech (Van Engen and Bradlow, 2007). Such a release in masking has also been demonstrated for other background speech languages such as Croatian (Calandruccio *et al.*, 2010) and Spanish (Garcia-Lecumberri and Cooke, 2006). Recently, two other studies have looked at the influence of background types other than native versus foreign or unfamiliar languages. For example, Russo and Pichora-Fuller (2008) found that younger listeners perform better when the background is familiar music than when it is unfamiliar music or babble, but for older listeners no differences between the background types were found. The authors suggested that the younger listeners "tuned into" the background music, whereas the older listeners "tuned out" the music. This increased attention seemed to be advantageous for the younger listeners. Boulenger *et al.*, (2010) showed how the token frequency of words (high vs low) that composed the babble influences target recognition. The results revealed that high-frequency babble interfered more strongly than low-frequency babble, indicating maximal competition from high frequency words.

In contrast to the target-masker language (mis)match effect found by the studies mentioned above, Mattys and colleagues (Mattys *et al.*, 2009; Mattys *et al.*, 2010) found no "language interference effect" in their extensive examinations of energetic and informational masking on speech segmentation in two-word phrases such as "mild option" versus "mile doption." In their (2009) study, they found no evidence that

the intelligibility of the masker affected listeners' segmentation patterns. That is, an intelligible masker (e.g., an English sentence) was not more distracting than its speech-shaped noise equivalent. Similarly, Mattys *et al.* (2010) found no evidence that the intelligibility of the competing talker (L1, L2, or an unintelligible language) had any effect on native English or native Cantonese listeners' responses to the segmentation of the two-word phrases. Mattys and colleagues addressed a number of stimulus and design differences between their work and our work (Calandruccio *et al.*, 2010; Van Engen and Bradlow, 2007) as an explanation for the inconsistent findings. Perhaps the most important of these differences is the fact that the relevant level of linguistic structure in the target and masker signals was different in their work but similar in ours. Specifically, in the speech segmental task used by Mattys and colleagues the targets were two-word phrases and the task demanded particular attention to the word juncture within a closed-set response format; whereas maskers were connected meaningful utterances taken from short story extracts. In our work, targets and maskers were both sentences and participants were required to report open set sentence recognition. Thus, as suggested by Mattys *et al.* (2010), their design features "…could have helped listeners maintain a high level of attention to task-relevant characteristics of the phrases and segregate the two streams effectively, keeping interference to a minimum" (page 11). In general, then, it appears that a background language effect, should it exist as an independent contributor to masking effectiveness, is likely to be constrained by task-related attentional factors. The background language effect might thus only appear when both targets and the maskers tap into the same level of processing (e.g., the sentence level). In the present study, we address this issue further by including a semantic content manipulation in addition to the target-masker language (mis)match manipulation. In this way, we can establish what happens when targets and maskers mismatch at the sentence level (i.e., meaningful target sentences embedded in semantically anomalous masker sentences).

A small number of studies have examined how listeners deal with target-masker language pairs that share many acoustic-phonetic characteristics some of which also involved a semantic content manipulation. For example, Freyman *et al.*, (2001) report that Dutch-accented English was a stronger masker than Dutch background speech for monolingual English listeners when presented with English targets consisting of semantically anomalous sentences. Tun *et al.*, (2002) found that older, but not younger, listeners received a release in masking for an unknown background language (Dutch) relative to a known background speech (English), and that they were more distracted by maskers that contained meaningful sentences than those consisting of randomly ordered word strings. The younger listeners in this study were efficient at tuning out the background speech in favor of target sentence processing regardless of the language or grammaticality of the background speech. However, these studies (Freyman *et al.*, 2001; Tun *et al.*, 2002) may have under- (or possibly over-) estimated the effects of background language and semantic content because they involved target stimuli that were either semantically anomalous (Freyman *et al.*, 2001) or were very long (20 words) (Tun *et al.*, 2002). These features of the target speech likely imposed a relatively high demand on processing resources which may have prevented relatively subtle effects of linguistic variation in the background speech from revealing themselves. In the current study, we minimize the memory component by using simple, meaningful target sentences (<7 words).

There is a growing interest in studying the effects of background language on second language recognition by bilingual listeners because of the well-known but not fully understood phenomenon of disproportionately detrimental effects of noise on second language relative to first language speech recognition (e.g., Mayo *et al.*, 1997; Nábêlek and Donahue, 1984; Rogers *et al.*, 2006, but see Cutler *et al.*, 2004 for equivalent effects of noise on a first and second language phoneme recognition task). This non-native language deficit appears to be modulated to some extent by the language of the background noise. Garcia-Lecumberri and Cooke (2006), for instance, showed that, when presented with English VCV stimuli, native Spanish listeners were overall less accurate at identifying the intervocalic consonant in all noise conditions (speech-shaped noise, multi-talker babble, and one-talker competing speech) compared to native English listeners. Moreover, while the English listeners benefited when the competing speech was in an unfamiliar language (Spanish) rather than in a familiar language (English), the Spanish-English bilingual listeners showed similar performance across English and Spanish maskers. Similarly, Van Engen (2010) tested monolingual English listeners and non-native English listeners, whose first language is Mandarin, on English target sentences in competing English and Mandarin two-talker babble. Van Engen found that, while both listener groups performed worse on the English than on the Mandarin babble, the native English listeners received a greater release from masking in the Mandarin versus English babble than the non-native listeners indicating roles for both language familiarity and target-to-background language (mis)match in speech-in-speech recognition. In the present study, we replicate and extend the examination of the background language effect for bilingual listeners by examining speech-in-speech recognition with both non-native (Experiment 2) and native (Experiment 3) target speech by bilingual listeners.

In the research presented here, we further examined the influence of linguistic similarity between the target and the background speech in the task that appears most likely to be sensitive to variation in the linguistic features of the masker, namely open set recognition of short, meaningful sentences. In particular, in this study we selected background languages that are either identical to or closely related phonetically to the target language for comparison to our earlier work with identical versus distantly related target and masker languages (English vs Mandarin-Chinese; English vs Croatian). We also compared semantically anomalous versus meaningful sentences as maskers for meaningful target sentences, and examined performance by English monolinguals (Experiment 1) and by Dutch-English bilinguals presented with English target sentences (Experiment 2) and with Dutch target

TABLE I. Designs of Experiment 1, 2, and 3.

Listeners	Background language	Content type	L1 targets	L2 targets (unfamiliar)
English listeners	L1 background	meaningful	*Exp 1*	
	speech	anomalous	Exp 1	
	L2 background	meaningful	Exp 1	
	speech (unfamiliar)	anomalous	Exp 1	
Dutch listeners	L1 background	meaningful	*Exp 3*	Exp 2
	speech	anomalous	Exp 3	Exp 2
	L2 background	meaningful	Exp 3	*Exp 2*
	speech (familiar)	anomalous	Exp 3	Exp 2

Note: Italic cells indicate conditions in which the target and background speech language match.

sentences (Experiment 3). (See Table I for an overview of the experimental designs.) In keeping with the target-masker linguistic similarity hypothesis, we predicted that the closely related target-masker pairing (English and Dutch) would lead to a relatively small but significant background language effect such that English-in-Dutch (mismatched) is better recognized than English-in-English (matched) for both monolingual and bilingual listeners. We also predicted that sentences that trigger sentence-level semantic processing would interfere more with our meaningful target sentences than sentences that differ from the target sentences by being semantically anomalous (see Mattys *et al.*, 2010, for a similar suggestion).

II. EXPERIMENT 1

Experiment 1 examined speech-in-speech recognition by native English listeners when presented with meaningful English target sentences embedded in a background of two-talker English and Dutch babble, consisting of meaningful and semantically anomalous sentences.

A. Method

1. Participants

a. Listeners. Twenty monolingual American-English listeners participated, including 6 males and 14 females (mean age: 20 yrs, 2 months) from the undergraduate student body at Northwestern University in Evanston, IL. No listeners had any knowledge of Dutch. They filled out a questionnaire in which they reported having normal speech and hearing.[1] Participants also reported not having (a history of) learning disabilities and/or auditory processing disorders. All listeners were paid for their participation in this experiment.

b. Talkers. Five female voices were used to create the stimuli. The voices of three native speakers of American-English (one for the English target speech and two for the English masker speech) and the voices of two native speakers of Dutch (two for the Dutch masker speech) were recorded. The English speakers were graduate students in the Linguistics Department of Northwestern University and the Dutch speakers were graduate students at the Max Planck Institute for Psycholinguistics. We used female voices (age range = 25–33 yrs) for the English and the Dutch background speech stimuli in order to match the gender of the target speaker with the background speaker. This eliminated the possibility of using talker gender differences as a segregation cue thereby making this speech-in-speech intelligibility relatively difficult and more likely to reveal any contribution of linguistic factors than if gender were available as a segregation cue (Brungart *et al.*, 2001).

2. Materials

a. Target and background speech sentences. English target sentences were taken from the revised Bamford-Kowal-Bench (BKB-R) Standard Sentence Test (Bamford and Wilson, 1979; Bench *et al.*, 1979). We chose the BKB-R as our target sentences because the sentences are syntactically simple and they include words that have been shown (Bent and Bradlow, 2003) to be highly familiar to nonnatives (Experiment 2). From a total of 21 BKB-R lists, we selected 8 lists (1, 7, 8, 9, 10, 15, 19, and 21) on the basis of their equivalent intelligibility scores for normal-hearing children (Bamford and Wilson, 1979). Each list contained 16 simple, meaningful sentences with 3 or 4 keywords for a total of 50 keywords per list (e.g., *HE PLAYED with his TRAIN*; *The CAT is SITTING ON the BED*). In total, 128 sentences were presented containing 400 keywords (shown in capital letters in the examples above).

As background speech sentences, we used 200 English meaningful sentences from the Harvard/IEEE sentence lists (IEEE, 1969; e.g., *Rice is often served in round bowls*) and 200 English semantically anomalous sentences from the syntactically normal sentence test (SNST; Nye and Gaitenby, 1974; e.g., *The great car met the milk*). The English background speech sentences were translated into Dutch by a native Dutch speaker and checked by two other native speakers of Dutch. The background speech always consisted of two-talker babble because previous research has shown that strong informational masking effects are observed when the background noise consists of two rather than more competing talkers (Brungart *et al.*, 2001; Calandruccio *et al.*, 2010; Freyman *et al.*, 2004; Van Engen and Bradlow, 2007).

b. Recordings. The speakers were instructed to produce the sentences in a "conversational" style of speech. The English materials were recorded in a sound-attenuating double-walled booth at Northwestern University in Evanston, IL. The speakers read the sentences in a self-paced manner

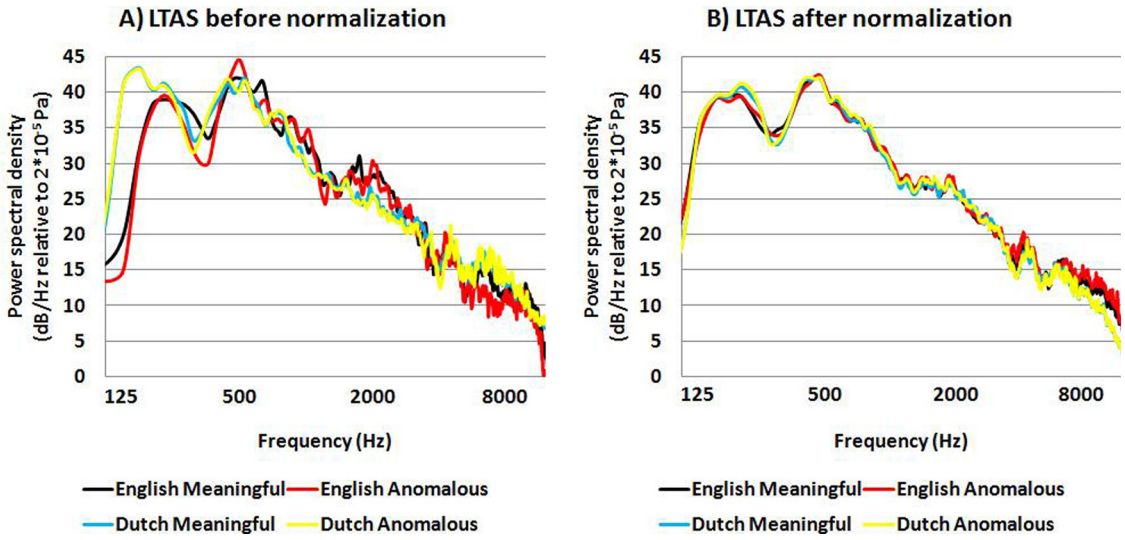

FIG. 1. (Color online) The LTAS of the four two-talker background speech maskers before (a) and after (b) the normalization.

from a computer screen. They spoke into a Shure SM81 Condenser microphone, and the sentences were recorded directly to disk using a MOTU Ultralite-mk3 external audio interface. The recordings were digitized at a sampling rate of 22050 Hz with 24 bit accuracy. The Dutch materials were recorded in a sound-attenuating booth at the Max Planck Institute for Psycholinguistics in the Netherlands. The speakers read all sentences from a written text. Their speech was recorded directly to a computer (sampling rate at 22050 Hz, 24 bit accuracy). The rms levels of all target and background sentences were equalized to the same pressure level.

c. Creating background speech tracks. To create the two-talker background speech tracks from the background sentence recordings, we did the following: For each of the four background speech talkers, we selected 100 meaningful and 100 anomalous sentences of their native language. We created eight different one-talker tracks by concatenating the files of each talker for each type (anomalous vs meaningful) in Praat (Boersma, 2001). In Audacity©, four different two-talker tracks were generated by mixing the talkers of the same language and the same content type. Speech was removed from the end of the track if it did not contain both talkers (due to differences in the duration of the two talkers' tracks). The four tracks were then equalized to the same rms level. The custom software our lab developed to run the experimental program then used the pressure level of each sound file to set the output level to 65 dB SPL as calibrated using a sound level meter attached to a Zwislocki coupler. Thus, during presentation, the overall level of the target sentences was fixed at 65 dB SPL, and the intensity of the two-talker babble background speech was varied to achieve the desired SNR levels. Specifically, for Experiment 1, background speech tracks were played at 68 dB SPL and 70 dB SPL to produce SNRs of −3 and −5 dB when mixed with the target sentences.

Before we mixed the background speech tracks with the target sentences, we manipulated the LTAS of all two-talker babble tracks as a means of reducing unequal amounts of energetic masking between the conditions. Figure 1 plots the LTAS of all four maskers before (a) and after (b) the normalization. Figure 1(a) shows substantial spectral differences in the higher frequencies (4 500–10 000 Hz) between the maskers. LTAS normalization eliminated these difference by adjusting each masker LTAS to match the average LTAS. The LTAS normalization procedure (implemented in MATLAB©, code available upon request) involved first computing the LTAS separately for each masker speech wave file (as shown in Fig. 1(a). The LTAS for a given wave file was computed by breaking up the file into windows of 2048 samples. The fast fourier transformation was then taken of each window and the mean was subsequently taken across all windows. After that, the average LTAS across all masker files was computed and each masker file LTAS was adjusted to the average LTAS.[2] Following this manipulation, we performed informal listening tests with native English and native Dutch lab members as listeners on the original and the spectrally-transformed sound files to ensure that the stimuli maintained their naturalness after signal processing. The results of these tests showed that normal-hearing listeners could not reliably distinguish between the original and normalized sound files. This was not surprising since the amount of spectral manipulation was very small.

Finally, the target sentences were mixed online with the LTAS-normalized two-talker background speech tracks using custom-designed software developed in Max/MSP (Cycling '74). The minimum length of the four background speech tracks was 2 min and 52 s. On each trial, a random portion of the desired two-talker background speech track was selected. A random number generator within our software was used to pick a point between 0 and 168 seconds. This number was used as the starting point within the background speech. The program held the time of each target sound file and played a portion of the background speech 1000 ms longer than the target. The two-talker background speech started 500 ms before the target sentence and continued for 500 ms after it.

3. Procedure

Listeners were tested individually, seated in a sound-attenuating booth. Participants received oral instructions. They were instructed to listen to English sentences spoken by a native English female speaker in the presence of two-talker background speech in their native language or a foreign/unfamiliar language. They were asked to repeat what they heard and were requested to report individual words if they were not able to identify the whole sentence. They could only listen to a sentence once. After a listener's response, the experimenter scored the response online and initiated the next trial. The responses were also recorded using an Olympus© digital voice recorder for subsequent reliability checking.

Stimuli were presented diotically with a MOTU 828 MkII input/output firewire device for digital-to-analog conversion (44100 Hz, 24 bit) passed through a Behringer Pro XL headphone amplifier, and output to MB Quart 13.01HX drivers. Participants wore disposable 13 mm foam insert earphones (Etymotic®). They started with eight practice trials (from list 20 of the BKB-R) to familiarize themselves with the task and the target talker. Half of the practice trials were presented at an SNR of +5 dB and the other half at an SNR of 0 dB. All four two-talker background speech types were randomly selected for these practice trials. After the practice session, participants were presented with a total of 128 experimental items (8 blocks of 16 sentences each). Trials were blocked by SNR level. We presented the experimental trials in the easy SNR level (−3 dB) before the experimental trials of the difficult SNR level (−5 dB) so that listeners could maximally adjust to the task and the target talker before being presented with the more challenging SNR. At each SNR, listeners were presented with a block of 16 sentences in each of the four background speech types (English Meaningful, English Anomalous, Dutch Meaningful, and Dutch Anomalous). Order of presentation of the four background speech types within each SNR block was randomized. The total duration of the experimental session was about 30 min.

4. Design and analysis

Listeners' recognition of the target sentences was based on their oral responses to three or four keywords per sentence. Two native listeners checked the reliability of the experimenter's (a non-native English listener, SB) online judgments. The inter-rater reliability was 94%. Each block of 16 sentences contained 50 keywords. Data were analyzed using linear mixed effects models (LMER, Baayen, Davidson, and Bates, 2008) with keyword identification (i.e., correct or incorrect) as the dichotomous dependent variable and Background Language (Match, e.g., English-in-English vs Mismatch, e.g., English-in-Dutch), Content Type (Anomalous vs Meaningful), and SNR (easy: −3 dB vs hard: −5 dB) as fixed factors. We used a logistic linking function to deal with the categorical nature of the dependent variable ([0,1]; cf., Dixon, 2008). Participant and item were entered as random effects in the model. In LMER models, a significant effect for a given factor can be inferred if the regression weight (beta) is statistically different from zero. Multiplication of the factor's beta by its numerical value gives the intercept adjustment associated with that factor. The sign of the weight indicates the direction of the deviation from the intercept. All binary predictor variables were contrast coded (i.e., −0.5 and +0.5, cf., Barr, 2008). This entails that the coefficient for each predictor represents the "main effect" for that predictor (i.e., its partial effect when all others are zero). In our analyses, we assigned the background speech that mismatched with the target language (i.e., Dutch in Experiment 1), anomalous background speech, and the easy SNR condition a negative weight (−0.5), whereas the background speech that matched with the target language (i.e., English), meaningful background speech, and the hard SNR condition were assigned a positive weight (+0.5). With this contrast coding, a negative regression weight would mean for the variable SNR (easy SNR = −0.5 vs hard SNR = +0.5) that the dependent variable has a higher value for the easy SNR than the hard SNR condition, i.e., participants performed better in the easy than in the hard SNR condition.

B. Results

The performance of the English listeners on the English target sentences is shown in Fig. 2 for (a) the easier SNR condition and (b) the harder SNR condition. The analysis on keyword identification revealed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.81$, $p < 0.0001$). The negative regression weight indicates that listeners performed better when the background language mismatched with the target language (Dutch) than when the background language matched with the target language (English). The analysis also showed a main effect of Content Type ($\beta_{ContentType} = -0.21$, $p < 0.0001$) and of SNR ($\beta_{snr} = -0.61$, $p < 0.0001$). The negative betas indicate that performance was better in the anomalous and easy SNR conditions versus the meaningful and hard SNR conditions. Finally, the interaction between Background Language and Content Type was significant ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.30, p < 0.01$).

To further investigate this interaction, we analyzed the effect of content type in each background language separately. The analysis showed that the content type effect was significant in the matched background language condition only ($\beta_{ContentType} = -0.07$, $p < 0.001$), indicating that the English listeners received a release from masking when the English background speech contained anomalous sentences as opposed to meaningful sentences. The content type effect was, however, not significant in the mismatched background language ($\beta_{ContentType} = -0.009, p > 0.1$). Thus, the interaction showed that listeners received a release from masking in anomalous English background speech versus meaningful English background speech, but not in anomalous Dutch versus meaningful Dutch.

C. Discussion

Experiment 1 showed, as expected, that higher SNRs (−3 dB) resulted in better target sentence recognition in all conditions (e.g., Brungart, 2001; Brungart *et al.*, 2001; Cooke *et al.*, 2008). Moreover, we found an effect of content type in the English background speech maskers only. That is, meaningful English maskers disrupted the recognition of

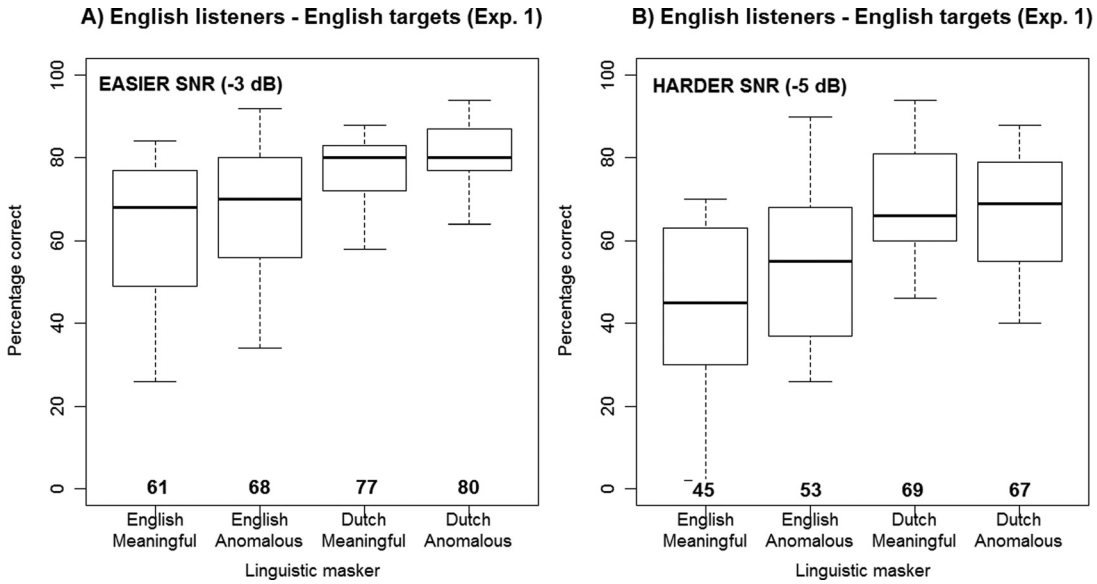

FIG. 2. Boxplots showing the interquartile ranges of intelligibility scores (in % correct) for English listeners on English target sentence recognition in (a) the easier SNR condition and in (b) the harder SNR condition (Experiment 1). Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. The mean is given at the bottom of each plot.

target speech more than anomalous English maskers. This result indicates that speech-on-speech masking not only involves interference at the phonetic/phonemic level, but also at the semantic level (cf., Tun *et al.*, 2002, for older listeners). The semantic content effect was, as expected, not present with the Dutch maskers because English listeners cannot understand the Dutch language. Hence, the listeners were not affected by the difference between meaningful and anomalous Dutch background speech.

In addition, we found that listeners received a release from masking when the competing speech was spoken in an unfamiliar language compared to their native language. This result is in line with previous findings involving speech-in-speech recognition (e.g., Calandruccio *et al.*, 2010; Garcia-Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007). The current results, however, cannot reveal whether the acoustic-phonetic similarity between the target (English) and masker language (Dutch) decreased the magnitude of the release relative to more phonologically unrelated language pairs (e.g., English-in-English vs English-in-Mandarin) since typological similarity was not manipulated under controlled experimental conditions within the same experiment. Nevertheless, this result suggests at a minimum that some degree of background-language-related release from masking is present even for a background language that is typologically very similar to the target language. In a companion paper, we report work from our laboratory that investigated English sentence recognition under three background language conditions within the same experiment (i.e., English, Dutch, and Mandarin-Chinese) by three different listener groups (i.e., native speakers of English, Dutch, and Mandarin-Chinese; see Calandruccio *et al.*, 2009, for a preliminary report).

III. EXPERIMENT 2

In Experiment 2, we examined how Dutch-English bilinguals—who are highly proficient in English and can thus understand information from both languages—perform on the same task and materials as in Experiment 1. The case of bilingual listeners allows us to manipulate target-masker linguistic similarity both with respect to the signals (English-in-English vs English-in-Dutch) and with respect to the listeners' linguistic knowledge and experience (native vs. non-native language targets and maskers). Moreover, we could test whether the (mis)match between the background language and the target language also plays an important role when bilinguals listen to targets in their second language (L2).

A. Method

1. Participants

a. Listeners. We selected 20 Dutch native participants from the Max Planck Institute's subject pool. This listener group included 3 males and 17 females ($M_{age} = 21$ yrs, 1 month). All listeners had completed their school education in the Netherlands, involving on average ten years of English lessons starting at age 11. The quantity and quality of exposure to English in the Netherlands is typically quite high therefore these Dutch-English bilinguals can be considered to be relatively high proficiency, non-native English listeners. Listeners reported no speech or hearing problems. They also reported not having (a history of) learning disabilities and/or auditory processing disorders. They were paid for their participation.

b. Talkers. The same talkers were used as in Experiment 1.

2. Materials

The same materials were used as in Experiment 1 except that we adjusted the SNR levels by 2 dB (i.e., −1 and −3 dB rather than −3 and −5 dB as in Experiment 1) for our non-native listener group. Previous work showed that even early

bilinguals—who learned their second language before age 6—were more detrimentally affected by noise in tasks of word or sentence recognition than monolinguals (Mayo *et al.*, 1997; Rogers *et al.*, 2006). In previous work in our laboratory with substantially lower proficiency non-native listeners than in the present study (e.g., Bradlow and Alexander, 2007) a +4 dB SNR adjustment for non-native relative to native listeners was sufficient to bring the two groups of listeners to the same level of baseline performance along the speech recognition accuracy scale. In the present study, we selected a +2 dB adjustment so that the two groups would have one common SNR (−3 dB) and there would be a 4 dB span between the hard SNR (−5 dB) for natives and the easy SNR for non-natives (−1 dB), thereby maximizing our chances of reaching the overall goal of similar baseline performance across the groups along the speech recognition accuracy scale. The background speech tracks were thus leveled at 66 and 68 dB SPL to produced SNRs of −1 and −3 dB when mixed with the English target sentences (presented at a fixed level of 65 dB SPL).

3. Procedure, design, and analysis

The procedure, design, and analysis were similar to the previous experiment except for some minor equipment-related differences. The Dutch listeners wore Sennheiser 280 Professional headphones during the experiment instead of insert ear phones. The experiment was run on a Windows computer and the stimuli were passed through an M-Audio fast Track Pro Audio/MIDI interface with preamps (96000 Hz, 24 bit).

B. Results

The performance of the Dutch listeners on the English target sentences is illustrated in Fig. 3 for (a) the easier SNR condition and (b) the harder SNR condition. The analysis on keyword identification showed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.24$, $p < 0.0001$). As indicated by the negative regression weight, Dutch listeners performed better when the competing speech mismatched (Dutch) than matched (English) with the target language. This shows that the (mis)match between the target and background language plays an important role, even when the target language is in the listener's L2. Further, the analysis revealed a main effect of Content Type ($\beta_{ContentType} = -0.13$, $p < 0.01$), with the negative regression weight indicating that listeners received a release from masking in anomalous versus meaningful maskers. The analysis also showed a main effect of SNR ($\beta_{snr} = -0.51$, $p < 0.0001$). The negative beta indicates that listeners performed better in the easy than in the hard SNR condition. The only significant interaction was between Background Language and Content Type ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.31\ p < 0.001$).

We further investigated this interaction by analyzing the effect of Content Type in each background language separately. The analysis showed that the content type effect was only significant in the matched background language condition ($\beta_{ContentType} = -0.07$, $p < 0.001$), but not in the mismatched background language condition ($\beta_{ContentType} = 0.007$, $p > 0.5$). This shows that Dutch listeners received a release from masking when the competing speech was anomalous English versus meaningful English, but not when the competing speech was anomalous Dutch versus meaningful Dutch despite their native familiarity with Dutch.

We next compared the performance of the English listeners (Experiment 1) with the Dutch listeners (Experiment 2). In our first cross-experiment analysis, we examined whether the 2 dB adjustment was effective at bringing the native and non-native listeners into the same range along the recognition accuracy scale. If this was the case, we would expect to find no main effect of Listener Group in the combined analysis. In the LMER analysis, we applied contrast-coding such that Dutch listeners were assigned a negative weight (−0.5) and

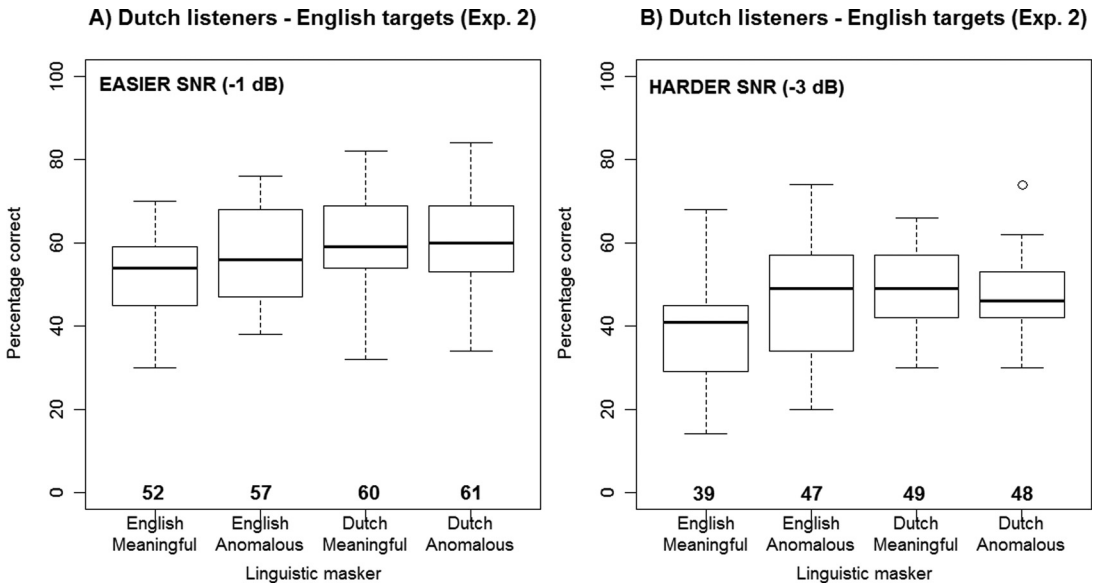

FIG. 3. Boxplots showing the interquartile ranges of intelligibility scores (in % correct) for Dutch listeners on English target sentence recognition in (a) the easier SNR condition and in B) the harder SNR condition (Experiment 2). Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. The mean is given at the bottom of each plot.

English listeners a positive weight (+0.5). The analysis showed a main effect of SNR ($\beta_{snr} = -0.55$, $p < 0.0001$), Background Language ($\beta_{BackgroundLanguage} = -0.42$, p 0.001), Content Type ($\beta_{ContentType} = -0.26$, $p < 0.05$), and Listener Group ($\beta_{ListenerGroup} = 0.79$, $p < 0.0001$). The positive regression weight for Listener Group indicates that the English listeners performed better overall than the Dutch listeners suggesting that, contrary to our expectation, the 2 dB adjustment was not sufficient to bring the two listener groups into the same performance range. Moreover, there was a three-way interaction between SNR, Background Language, and Content Type ($\beta_{SNR\ x\ BackgroundLanguage\ x\ ContentType} = -0.28$, $p < 0.05$). Separate analyses for the easy SNR (−3 dB for native listeners, −1 dB for non-native listeners) and the hard SNR (−5 dB for native listeners, −3 dB for non-native listeners) showed that for both natives and non-natives, the Background Language by Content Type interaction was significant in the hard SNR ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.46\ p <$ 0.001) but not the easier SNR. In other words, in the hard SNR, both natives and non-natives received a release from masking when the competing speech was anomalous English versus meaningful English, but not when the competing speech was anomalous Dutch versus meaningful Dutch. This suggests that linguistic differences across speech maskers exert their effect most effectively under highly adverse conditions. We also note here that the interaction with SNR was not observed in the separate analyses for Experiments 1 and 2 probably due to a lack of statistical power; only when both groups of participants were included did three-way interaction reach statistical significance.

Given that a 2 dB adjustment was apparently not sufficient to equate overall performance by the native and non-native listeners, we next asked whether a 4 dB adjustment (cf., Bradlow and Alexander, 2007) would eliminate the main effect of Listener Group and in so doing allow us to directly compare native and non-native listeners in the same range of overall sentence recognition accuracy. We therefore then compared performance on the hard SNR level for the English listeners (i.e., −5 dB) versus performance on the easy SNR level for the Dutch listeners (i.e., −1 dB). Note that in this analysis SNR is removed as a factor because we are only looking at one SNR level (i.e., −5 dB for the natives and −1 dB for the non-natives) and that this analysis contains half the data of the first analysis since data from only one SNR were included from each listener group. The results showed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.55$, $p <$ 0.0001), Content Type ($\beta_{ContentType} = -0.12$, $p < 0.05$), and an interaction between Background Language and Content Type ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.31$, $p < 0.01$). The interaction revealed that both listener groups received a release from masking when the competing speech was anomalous English versus meaningful English, but not when the competing speech was anomalous Dutch versus meaningful Dutch. Importantly, the analysis revealed no main effect of Listener Group ($\beta_{ListenerGroup} = 0.06$, $p > 0.1$), indicating that English and Dutch listeners performed at similar average accuracy levels when the SNR level was raised by 4 dB rather than by 2 dB for the non-native listeners relative to the native listeners. This also suggests that Dutch listeners are, perhaps unexpectedly, more similar to the low proficiency non-native listeners in our previous work (cf., Bradlow and Alexander, 2007) in that they need a 4 dB adjustment to get into the same level of accuracy as the native listeners. However, we still found an interaction effect between Background Language and Listener Group ($\beta_{BackgroundLanguage\ x\ ListenerGroup} = -0.61$, $p < 0.0001$), indicating that Dutch listeners received a smaller release from masking when the background language mismatched (Dutch) versus matched (English) with the target language than English listeners.

Finally, in a third cross-experiment analysis, we looked at the one SNR that the two listener groups have in common (i.e., −3 dB). This comparison allowed us to compare performance across the two listener groups under identical signal conditions. SNR is again removed as a factor because we are only looking at one SNR, and this analysis again contained half the data of the first cross-experiment analysis since data from only one SNR were included from each listener group. The analysis showed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.49$, $p < 0.001$), Content Type ($\beta_{ContentType} = -0.21$, $p < 0.0001$), and Listener Group ($\beta_{ListenerGroup} = 1.19$, $p < 0.0001$). As expected, the main effect of Listener Group indicates that when identical cues are presented to natives and non-natives, they do not perform at a similar average level pointing to the need to adjust the SNR level of non-native listeners when seeking to identify significant predictors of deviation from the average level of sentence recognition accuracy. As expected from the previous analyses, this analysis also revealed an interaction between Background Language and Content Type ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.30$, $p < 0.01$) and between Background Language and Listener Group ($\beta_{BackgroundLanguage\ x\ ListenerGroup} = -0.51$, $p < 0.0001$). The first interaction indicates that both listener groups received a release from masking when the competing speech was anomalous English versus meaningful English, but not when the competing speech was anomalous Dutch versus meaningful Dutch. The second interaction indicates that Dutch listeners received a smaller release from masking when the background language mismatched (Dutch) versus matched (English) with the target language than English listeners.

C. Discussion

Experiment 2 examined how Dutch-English bilinguals perform on a speech-in-speech recognition task with English target sentences. The results showed that Dutch listeners received a release from masking when the competing speech is different from the target speech even though the competing speech is in their L1 (Dutch) and the target speech is in their L2 (English). Dutch background speech (native language) is thus less disruptive than English background speech (foreign, but familiar language) when listening to English targets, indicating that the language match between the target and the background speech is more interfering than listener's familiarity with the language of the background speech. Note, however, that familiarity with the background language also plays a role because the cross-experiment analysis showed that the release from masking

when the competing speech was Dutch versus English was smaller for Dutch listeners than English listeners (see Van Engen, 2010, for a similar finding with Mandarin instead of Dutch background speech and listeners).

The Dutch listeners' ability to suppress processing of the Dutch background speech is also supported by the finding that the semantic content effect only appeared with the English and not with the Dutch background speech. That is, Dutch listeners performed worse on meaningful English versus anomalous English background speech, but no such differences were found between the Dutch background speech conditions despite the fact that these native Dutch listeners could surely easily distinguish meaningful from semantically anomalous Dutch sentences. This suggests that sensitivity to sentence-level semantic content in the masker speech is only high under conditions of target-masker language identity (in this case, English-in-English). It thus appears that these Dutch-English bilingual listeners were able to suppress native language sentence-level semantic processing when listening to target speech in their second language. Note, however, that this mechanism may only work for very proficient bilingual listeners such as the Dutch listeners in this study as opposed to the Mandarin listeners in Van Engen's (2010) study (although note that Van Engen did not investigate the effect of variation in the semantic content of the masking speech).

IV. EXPERIMENT 3

Experiment 3 extended the investigation by examining how Dutch listeners perform on targets in their native language, while using the same background speech materials as in Experiment 1 and 2. In this way, we could establish whether background language semantic processing is also suppressed when listening to native language targets. In addition, as far as we know, this is the first test of native rather than non-native target recognition when the targets are presented mixed with both L1 and L2 background speech.

A. Method

1. Participants

a. Listeners. We selected 20 native Dutch participants recruited from the same population as the participants in Experiment 2 (i.e., the Max Planck Institute's subject pool). None had taken part in Experiment 2. This listener group included 2 males and 18 females ($M_{age} = 21$ yrs, 6 months). All listeners started English lessons at age 11 and therefore had on average ten years of education in English. No speech or hearing problems were reported. They also reported not having (a history of) learning disabilities and/or auditory processing disorders. They were paid for their participation.

b. Talkers. For the Dutch targets we used the voice of a female native Dutch speaker (postdoctoral researcher at the Max Planck Institute for Psycholinguistics). The English target sentences were excluded from this experiment.

2. Materials

The same background speech materials were used as in Experiments 1 and 2. Note that we used the same SNR levels as in Experiment 1 (−3 dB and −5 dB) because the listeners were listening to their native language. We used the target speech from the Dutch speaker. The Dutch target sentences were direct translations of the English BKB-R sentences. One native Dutch speaker translated all sentences which were checked by two other native speakers of Dutch.

The Dutch target sentences were recorded in a sound-attenuating booth at the Max Planck Institute for Psycholinguistics in the Netherlands. The speaker read all sentences from a written text. Her speech was recorded directly to a computer (sampling rate at 22050 Hz).

3. Procedure, design, and analysis

The same procedure, design, and analysis were used as in Experiments 1 and 2. In the LMER analysis, we again applied contrast-coding for all the fixed factors. Note, however, that the target language in this experiment changed from English to Dutch compared to the previous experiments. As a result, we assigned English, the mismatched background language, a negative regression weight (−0.5) and Dutch, the matched background language, a positive regression weight (+0.5).

B. Results

Figure 4 shows the performance of the Dutch listeners on the Dutch target sentences for (a) the easier SNR condition and (b) the harder SNR condition. The analysis on keyword identification revealed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.63$, $p < 0.0001$). The negative beta shows that Dutch listeners received a release from the background language that mismatched (English) versus the background language that matched (Dutch) when listening to Dutch. Moreover, we found a main effect of Content Type ($\beta_{ContentType} = -0.10$, $p < 0.05$). The negative regression weight indicates that Dutch listeners performed better on anomalous versus meaningful background sentences. Finally, the analysis showed a main effect of SNR ($\beta_{snr} = -0.53$, $p < 0.0001$), indicating that listeners performed better in the easy than the hard SNR condition. There were no significant interaction effects (all p's > 0.05).

We also compared the performance of the English listeners on English target speech (Experiment 1) with the performance of the Dutch listeners on Dutch target speech (Experiment 3). In the LMER analysis, we applied contrast-coding such that Dutch listeners were assigned a negative weight (−0.5) and English listeners a positive weight (+0.5). Note also that we assigned a negative weight (−0.5) to the background language that mismatched with the target language (i.e., Dutch in Experiment 1 and English in Experiment 3) and a positive weight (+0.5) to the background language that matched with the target language (i.e., English in Experiment 1 and Dutch in Experiment 3). The analysis showed a main effect of SNR ($\beta_{snr} = -0.56$, $p < 0.0001$), Background Language ($\beta_{BackgroundLanguage} = -0.49$, p

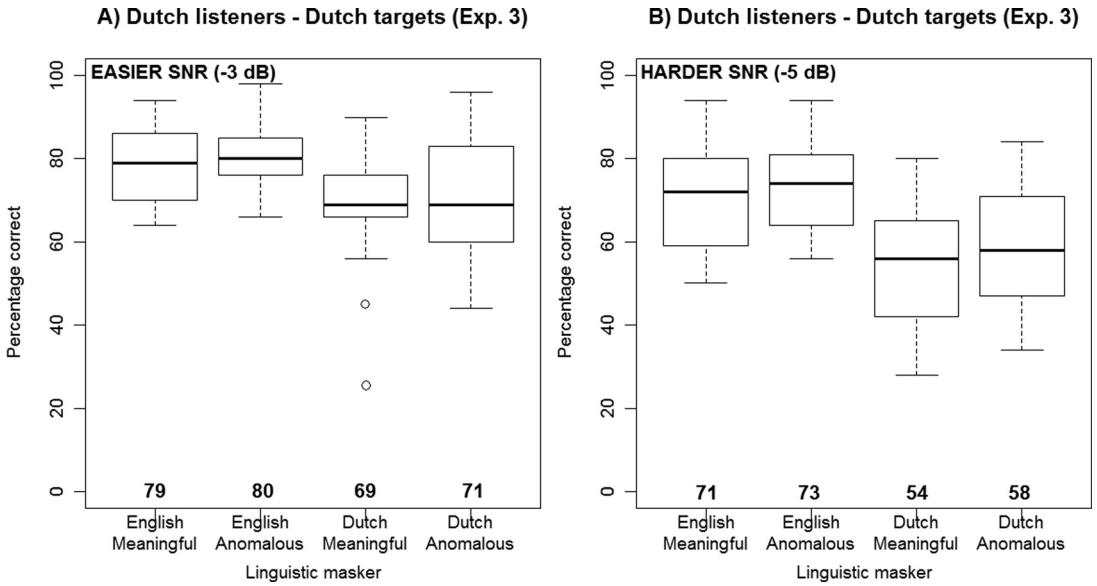

FIG. 4. Boxplots showing the interquartile ranges of intelligibility scores (in % correct) for Dutch listeners on Dutch target sentence recognition in (a) the easier SNR condition and in (b) the harder SNR condition (Experiment 3). Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. The mean is given at the bottom of each plot.

< 0.0001), and Content Type ($\beta_{ContentType} = -0.23$, $p < 0.05$). The lack of a main effect of Listener Group shows that the English listeners performed overall similar to the Dutch listeners. (Recall that both of these native listener groups received the same SNR levels of -3 dB and -5 dB for the easy and hard conditions, respectively). In addition, no factors interacted with Listener Group (all p's ≥ 0.05), indicating that the status of the listeners (native vs non-native) with respect to the background language does not influence recognition accuracy when listening to native language targets. This also confirms that the Dutch and English target stimuli were of about equivalent inherent intelligibility to native listeners. There was a significant interaction between SNR and Background Language ($\beta_{snr\ x\ BackgroundLanguage} = -0.15$, $p < 0.05$), arising from the larger matched-to-mismatched background language difference in the hard relative to the easy SNR in both Experiment 1 and Experiment 3. With this combined analysis there appears to be sufficient power for this interaction to reach statistical significance whereas the separate analyses for each experiment (half the number of participants) did not show this interaction. Although Background Language reached significance in both SNR conditions, the effect size in the harder SNR conditions (17.5%) was bigger than in the easier SNR conditions (11.5%). These effect sizes are noteworthy because it demonstrates that, despite the constant energetic masking differences between the matched and mismatched background language maskers across SNRs, a linguistic masking difference based on target-language (mis)match emerges more strongly under one (the more challenging listening condition) but not the other listening condition. Note however that it is also possible that this interaction was driven by ceiling effects in the easier SNR conditions.

Next, we compared the performance of Dutch listeners on English targets (Experiment 2, non-native language targets) versus Dutch targets (Experiment 3, native language targets). First we conducted this analysis on the full set of data, that is with data from both SNRs for both experiments (-3 and -1 dB for Experiment 2; -5 and -3 dB for Experiment 3). In the LMER analysis, we applied contrast-coding such that the Dutch listeners in Experiment 2 were assigned a negative weight (-0.5, English targets) and the Dutch listeners in Experiment 3 a positive weight ($+0.5$, Dutch targets). Note also that we assigned a negative weight (-0.5) to the background language that mismatched with the target language (i.e., Dutch in Experiment 2 and English in Experiment 3) and a positive weight ($+0.5$) to the background language that matched with the target language (i.e., English in Experiment 2 and Dutch in Experiment 3). The analysis showed a main effect of SNR ($\beta_{snr} = -0.51$, $p < 0.0001$), Background Language ($\beta_{BackgroundLanguage} = -0.43$, $p < 0.05$), Content Type ($\beta_{ContentType} = -0.11$, $p < 0.01$), and Target Language ($\beta_{TargetLanguage} = 0.81$, $p < 0.0001$). The positive regression weight for Target Language shows that the Dutch listeners performed better overall when the targets were in Dutch (Experiment 3) than in English (Experiment 2). This is consistent with Experiment 1 versus Experiment 2 analysis above which showed that the 2 dB adjustment for non-native versus native language targets was not sufficient to bring native and non-native recognition accuracy into the same range. We also found an interaction between Background Language and Content Type ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.16$, $p < 0.05$), between Background Language and Target Language ($\beta_{BackgroundLanguage\ x\ TargetLanguage} = -0.38$, $p < 0.0001$), and a three-way interaction between Background Language, Content Type, and Target Language ($\beta_{BackgroundLanguage\ x\ ContentType\ x\ TargetLanguage} = 0.28$, $p < 0.05$). The three-way interaction (Background Language, Content Type, and Target Language) indicates that the release from masking associated with a mismatched relative to a matched background language

was larger for native than non-native language listening, and this difference was especially marked in the anomalous background speech conditions.

Finally, we compared performance on the easy SNR in Experiment 2 (i.e., −1 dB for non-native language targets) versus performance on the hard SNR in Experiment 3 (i.e., −5 dB for native language targets) to examine whether a 4 dB adjustment would be enough to remove the main effect of Target Language. Note that in this analysis SNR is removed as a factor because we are only looking at one SNR level and only half of the data are included (only one SNR for each experiment). The results showed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.47$, $p < 0.0001$), Content Type ($\beta_{ContentType} = -0.12$, $p < 0.05$), and Target Language ($\beta_{TargetLanguage} = 0.30$, $p < 0.0001$). The positive regression weight of Target Language shows that, even when the SNR level is adjusted by 4 dB, Dutch listeners still perform better on their native than on a non-native language. We also found an interaction between Background Language and Target Language ($\beta_{BackgroundLanguage\ x\ TargetLanguage} = -0.46$, $p < 0.0001$), indicating that Dutch received a bigger release from masking for the mismatched versus the matched background language when listening to Dutch targets than to English targets. No other interactions were significant.[3]

C. Discussion

Experiment 3 showed that the Dutch-English bilingual listeners received a release from the mismatched (English) versus the matched (Dutch) background language when they performed a speech-in-speech recognition task with L1 targets (Dutch). The same pattern was found in Experiment 2, in which the bilinguals also received a release from the mismatched (Dutch) versus the matched (English) background language, however, the release was bigger in Experiment 3 with native language targets than in Experiment 2 with non-native language targets. This shows that the effect of background language on the recognition of speech by bilinguals not only depends on the (mis)match between target and masker, and on the listener's familiarity with the background language, but also depends on whether the task is a native or non-native language recognition task. This study replicates and extends previous findings by showing not only a release from masking for non-native listeners on L2 targets (e.g., Garcia-Lecumberri and Cooke, 2006; Van Engen, 2010), but also on L1 targets.

Surprisingly, in Experiment 3 with Dutch targets, we did not find an interaction between background speech language and content type. Based on the findings of Experiment 2, we expected that the Dutch listeners would receive a release from masking for anomalous versus meaningful background sentences only in the same language condition (i.e., when the background language matched the target language). In other words, we predicted that, when listening to Dutch targets, Dutch listeners would be sensitive to meaning in the Dutch background speech and insensitive to meaning in the English background speech. Possible other explanations for the lack of such a pattern are given in Sec. V.

V. GENERAL DISCUSSION

This study examined whether speech-in-speech recognition varies across phonetically close target and background languages (English vs Dutch), across variation in the semantic content of the masker (meaningful vs semantically anomalous sentences), as well as across listener status with respect to both the target and the background language (native vs non-native listeners). The overarching goal of this research was to test the target-masker linguistic similarity hypothesis, that is, to examine whether similarity in various aspects of linguistic information carried by the target and the masker signals contributes to speech-on-speech masking.

Experiment 1 presented English target sentences in English and Dutch background speech to monolingual English listeners. The results showed that listeners received a release from masking when the competing speech was spoken in an unfamiliar language versus their native language, replicating previous work (e.g., Calandruccio *et al.*, 2010; Garcia-Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007). The innovative aspect of this finding is that the background language release from masking was found for phonetically closely related languages (English vs. Dutch) as opposed to phonetically more distant languages (e.g., English vs Mandarin; English vs Spanish). This result is in line with the findings by Freyman *et al.* (2001) and Tun *et al.* (2002) who also looked at phonetically closely related language pairs, however, those studies used targets that were either semantically anomalous or were very long and complex. The current study presented simple, meaningful target sentences, thereby reducing the demands on target processing in order to highlight effects of variation in language and semantic content of the background sentences. Importantly, we also found an effect of content type in the English background speech maskers only. That is, listeners received a release from masking when the competing speech was anomalous English as opposed to meaningful English, whereas no such effect was found in the Dutch background speech conditions. This result indicates that linguistic masking in speech-in-speech recognition involves interference from relatively abstract levels of linguistic structure (i.e., sentence-level semantics).

Experiment 2 used the same materials as in Experiment 1, but tested Dutch-English bilinguals. The findings demonstrated that these listeners performed better on English targets when the competing speech was Dutch versus English. This result is in line with Van Engen (2010), but contrasts with previous findings by Garcia-Lecumberri and Cooke (2006). In their study, non-native Spanish listeners performed at a similar level irrespective of whether the masker language was in their first (Spanish) or second language (English). The authors suggest that L1 background speech might in general be more difficult to ignore than L2 background speech (linguistic familiarity), but that the task of recognizing L2 consonants may have made it harder to ignore L2 background speech (linguistic similarity). In this way, the differences between L1 and L2 background speech upon L2 target recognition may have cancelled each other out. There are three possible explanations for these

conflicting findings: differences in listener groups, in the type of task, and in the number of background speakers. The non-native listeners in our study were highly proficient bilinguals, whereas the non-native listeners in Garcia-Lecumberri and Cooke's (2006) study were less proficient in their second language (but note that the non-native listeners in Van Engen, 2010 also had relatively low proficiency in their second language). In addition, our task tapped into higher levels of processing (recognition of sentences) than their study (identification of phonemes). Finally, we used two-talker babble instead of one competing talker. A combination of these factors could have led to the different results in the two studies. Further studies with different groups of bilingual listeners (low vs. high proficiency), different materials (phonemes, words, and sentences), and different number of background talkers are therefore necessary.

Experiment 2 also found a significant interaction between background language and content type, indicating that the Dutch-English bilinguals were affected by the semantic content of the English maskers (8.5% increase in performance) but not by the semantic content of the Dutch maskers during L2 target recognition (0% decrease in performance). Bilingual listeners may thus be quite effectively inhibiting L1 processing such that competing speech in their native language is processed in a "shallow" way when attending to L2 speech targets. This idea is in line with results from a study by Colzato *et al.*, (2008). In their study, monolinguals and bilinguals performed a rapid serial visual presentation task, in which they were asked to report two digits (T1 and T2) presented in a stream of letter distractors. The lag between T1 and T2 varied randomly. Such a task could produce an attentional blink (Raymond *et al.*, 1992), which occurs when two masked target stimuli appear in close proximity. Generally, if T1 is correctly reported, people have difficulty reporting T2 when it occurs within an interval of about 100−500 ms after T1. Colzato *et al.* (2008) showed that bilinguals displayed a larger attentional blink than monolinguals, that is, a bigger decrease in performance at shorter lags. The authors suggested that "bilinguals invest more of their resources in processing a target and/or processing a target leads to a stronger inhibition of competitors" (p. 310). Consistent with this suggestion, our results show that when bilinguals focus on second language (English) speech recognition, their processing resources are primarily committed to relevant information (English speech) resulting in a reduction of processing resources available for competing irrelevant information (Dutch speech) (cf., Bialystok *et al.*, 2006).

Comparing the results of Experiment 1 and Experiment 2 showed that native and non-native listeners perform similarly overall when the SNR level was adjusted by 4 dB. This result indicates that we cannot directly assume that our non-natives were such high proficiency listeners in English. Future research should therefore take into account participants' English language experience such as their self-rated proficiency in English. However, Dutch listeners received a smaller release from masking than English listeners when the background language mismatched (Dutch) versus matched (English) the target language, indicating that familiarity with the target language plays a role. This replicates the results of Van Engen (2010) who showed the same pattern of effects with participants who also speak both background languages, i.e., native Mandarin speakers.

Experiment 3 presented Dutch target sentences in English and Dutch background speech to Dutch-English bilinguals. The critical difference between Experiment 1 and Experiment 3 was that in Experiment 3 the L2 background speech was familiar to the participants whereas in Experiment 1 the monolingual listeners were only familiar with the background language that matched the target language. The Dutch listeners performed better when the competing speech was spoken in their second language (English) versus their native language (Dutch). This is the first evidence that bilingual participants, who are listening to their native language, experience the same release from masking as monolinguals when the competing speech is in a foreign (and familiar) language versus in their native language. (Note that the bilinguals in Garcia-Lecumberri and Cooke (2006) and Van Engen (2010) were listening to L2 targets only).

The results of Experiment 3 showed, unexpectedly in view of the results from Experiments 1 and 2, no significant interaction between background language and content type. In Experiment 2 we found that listeners' performance increased with 8.5% on average for English meaningful as compared to English anomalous speech, whereas we found a −0.5% decrease in performance in Experiment 3. We predicted, as in Experiment 2, an effect of semantic content in the Dutch background speech condition (i.e., a significant difference between anomalous Dutch vs meaningful Dutch), but not in the English background speech condition (i.e., no difference between anomalous English and meaningful English) because their task was to recognize the target Dutch sentences. However, the results showed a main effect of semantic content, indicating that the Dutch listeners were sensitive to meaning independent of the background speech language in this experiment. There are three possible explanations for this finding. First, it is possible that there is not enough power to reveal the interaction between background language and content type in Experiment 3. Figure 4 shows that the anomalous versus meaningful effect is numerically smaller for the English background speech condition than for the Dutch background speech condition. However, since we found a significant interaction between background language and content type with the same number of subjects and items in Experiments 1 and 2 as in Experiment 3, lack of statistical power is an unlikely explanation for the lack of a background language by content interaction in Experiment 3.

An alternative possible explanation for the lack of a background language by content interaction in Experiment 3 is that spectro-temporal differences between the Dutch and English targets somehow caused the background language by content effect to be neutralized in the one case but not the other. We found that the LTAS curve of the Dutch target track had greater energy than the LTAS curve of the English target track in the higher frequencies, which may make the Dutch targets more robust against background speech. It could therefore be the case that the Dutch targets were very easily recognized by the native Dutch listeners leaving

plenty of processing resources available for semantic processing of the background speech in either language. However, the similar overall levels of performance demonstrated by the comparison of Experiment 1 (English targets and English native listeners) and Experiment 3 (Dutch targets and Dutch native listeners) contradicts this account.

A third possible explanation for the lack of a background language by content interaction is that listeners may allocate their resources differently depending on whether they listen to L1 (Experiment 3) or L2 speech targets (Experiment 2). In Experiment 2, the additional resources required for processing L2 speech targets led to a suppression of semantic processing in background L1 speech. Semantic processing of background L2 speech was not inhibited in this experiment as it matched the target language, and the meaningful target sentences would have engaged semantic processing of that language in both the target and background speech. However, semantic processing was independent of the background language when the target language changed from the bilinguals' L2 to their L1 (Experiment 3). In this case, the less resource demanding task of L1 target recognition did not involve the suppression of semantic processing in either background language. Evidence in favor of this processing resource account could be gathered by making the task in Experiment 2 easier, for example by using a higher SNR, such that fewer resources are needed during L2 processing. Conversely, the task in Experiment 3 could be made harder in order to increase the processing resources required for L1 recognition.

The processing resource account outlined above would be consistent with the notion of processing costs associated with switching between languages. The phenomenon of task switching costs in general cognitive tasks (e.g., Rogers and Monsell, 1995) has been explored for the particular case of language switching in bilinguals (e.g., Bialystok *et al.*, 2008; Meuter and Allport, 1999) and generally shows asymmetries in the suppression/activation profiles of L1 and L2 with some modulation by the relative proficiencies in each of the two languages (Costa and Santesteban, 2004) and by the onset age of bilingualism (Luk *et al.*, 2011). Interestingly, (and somewhat counter-intuitively) the asymmetry is with respect to difficulty of inhibition rather than ease of activation. Thus, for example, it is harder for bilinguals to switch from L2 to L1 in a speech production task than vice versa because it will take longer to switch into a language which is more suppressed, i.e., L1. In our sentence recognition task with either matched or mismatched-language target-masker pairings, we found that the bilingual listeners did not suppress mismatched-language semantic processing in the case of L1 targets (Experiment 3, Dutch-in-English) whereas they did suppress mismatched-language semantic processing in the case of L2 targets (Experiment 2, English-in-Dutch). Contrary to this pattern, the switch cost idea would seem to lead us to suppose that the Dutch background speech should be harder to suppress than English background speech. However, the present data point to the potential importance of the target language activation: L1 target activation somehow promotes general language-independent semantic processing of background speech while L2 target activation somehow promotes language-selection for semantic processing. Further exploration of this pattern is needed to fully understand this aspect of bilingual speech-in-speech processing as well as its connection to the now well-documented bilingual advantage in executive control functioning relative to monolinguals (for an overview, see Adesope *et al.*, 2010; Bialystok, 2007).

Comparing the results of Experiment 1 with Experiment 3 showed that the status of the listeners with respect to the languages in the background (native vs totally unfamiliar in Experiment 1, native vs non-native in Experiment 3) had no effect on how listeners process background speech when listening to native speech. The release from masking due to a target-background language mismatch for bilinguals, who were familiar with both background languages, was similar to the language mismatch masking release for monolinguals, who only knew one background language. In contrast, the comparisons between Experiment 1 and Experiment 2 (both with English targets), and between Experiment 2 and Experiment 3 (both with Dutch listeners) revealed that Dutch listeners received a bigger release from the mismatched background language versus the matched background language when listening to Dutch (Experiment 3) than English (Experiment 2) targets. This effect was most pronounced in the conditions with semantically anomalous background speech. This complex interplay of the target-masker language relation with listener language experience indicates that, even under conditions of a constant target-masker energetic masking difference, linguistic factors seem to play an important role in determining speech-in-speech recognition independently of more general, auditory, non-linguistic factors that underlie target-to-masker separation. Further supporting a role for linguistic processing as a dimension of stream segregation for speech-in-speech recognition is the three-way interaction between SNR, Background Language, and Content Type observed in the combined analysis of Experiments 1 and 2, and the significant interaction between SNR and Background Language in the combined analysis of Experiments 2 and 3. These interactions suggest that factors that can increase target-masker similarity, such as semantic content and language, have an especially strong effect on masking effectiveness when the listening conditions are highly unfavorable. This argues against the notion that the effects observed here could be attributed to energetic masking differences between the various speech maskers since those differences should remain constant across SNR conditions. Instead, it appears that when the auditory system is more taxed, we observe performance differences that are directly related to target-masker distance along linguistically defined dimensions (see Calandruccio *et al.*, 2010 for additional data and discussion of this point).

In conclusion, the present research has provided evidence in support of the hypothesis that target-masker linguistic similarity plays a significant and independent role in determining speech-in-speech recognition accuracy. The more the target speech matches the masker speech along linguistically defined dimensions (e.g., same vs different language, same vs. different levels of semantic content), the greater the interference on

speech recognition accuracy. Moreover, this study revealed a complex interplay between linguistic release from masking and the listener's knowledge of the language of the target and of the background language. Thus, while signal-bound, energetic masking differences may dominate stream segregation for speech-in-speech recognition, target-masker linguistic similarity likely makes an independent contribution raising the possibility of speech-in-speech enhancement strategies that focus on these factors.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01-DC005794 from NIH-NIDCD, and the Hugh Knowles Center at Northwestern University. We would like to thank Sumitrajit Dhar for the use of his equipment and Chun Liang Chan for his technical support throughout this project. We also thank Andrew Sabin for help with the LTAS normalization procedure. Parts of this work were presented at the Acoustical Society of America 2009 (Portland) and 2010 conference (Cancun).

[1]Throughout this study we limited participation to listeners who reported normal speech and hearing abilities. Although participants were not subjected to a hearing screening, the distribution of sentence recognition accuracy scores showed only two outliers ($M = 26$ and 44) in the easier meaningful Dutch condition (Fig. 4). These two participants scored lower than average in this condition. We looked at the performance of these participants in all the conditions to examine if they performed worse overall. We found that they mainly had difficulty recognizing target speech in the Dutch language conditions, but not in the English language conditions. This indicates that their lower performance in the Dutch language conditions compared to the rest of the participants could most likely not be explained by any hearing loss. This thus suggests that there were most likely no participants with hearing loss that would affect overall speech recognition performance.

[2]This adjustment was performed as follows. Let the LTAS of 1 masker be mLTAS (length is 2048, each point is an amplitude at a frequency). Let the grand average LTAS be gaLTAS (length is 2048, each point is an amplitude at a frequency). The adjusted masker LTAS is created by performing the following operation to each of the 2048 samples: adjustedLTAS(f) = mLTAS(f) * (gaLTAS(f)/mLTAS(f)) where f iterates across each frequency in the spectrum.

[3]For completeness, we also looked at the one SNR that the two listener groups have in common (i.e., −3 dB). This comparison allowed us to compare performance across the two listener groups under identical signal conditions. SNR is again removed as a factor because we are only looking at one SNR level. This analysis showed a main effect of Background Language ($\beta_{BackgroundLanguage} = -0.38$, $p < 0.0001$), Content Type ($\beta_{ContentType} = -0.11, p < 0.05$), and Target Language ($\beta_{TargetLanguage} = 1.33, p < 0.0001$). As expected, the main effect of Target Language indicates that when Dutch listeners could make use of identical cues, they do not perform similarly on their native versus a non-native language. Moreover, we found an interaction between Background Language and Content Type ($\beta_{BackgroundLanguage\ x\ ContentType} = -0.22$, $p < 0.05$), between Background Language and Target Language ($\beta_{BackgroundLanguage\ x\ TargetLanguage} = -0.29$, $p < 0.0001$), and between Background Language, Content Type, and Target Language ($\beta_{BackgroundLanguage\ x\ ContentType\ x\ TargetLanguage} = 0.46, p < 0.05$).

Adesope, O. O., Lavin, T., Thompson, T., and Ungerleider, C. (**2010**). "A systematic review and meta-analysis of the cognitive correlates of bilingualism," Rev. Educ. Res. **80**(2), 207–245.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Bamford, J., and Wilson, I., "Methodological considerations and practical aspects of the BKB sentence lists," in *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children*, edited by J. Bench and J. Bamford (Academic, London, 1979), pp. 148–187.

Barr, D. J. (**2008**). "Analyzing 'visual world' eye tracking data using multilevel logistic regression," J. Mem. Lang. **59**, 457–474.

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

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech and intelligibility benefit," J. Acoust. Soc. Am. **114**(3), 1600–1610.

Bialystok, E. (**2007**). "Cognitive effects of bilingualism: How linguistic experience leads to cognitive change," International Journal of Bilingual Education and Bilingualism **10**, 210–223.

Bialystok, E., Craik, F., and Luk, G. (**2008**). "Cognitive control and lexical access in younger and older bilinguals," J. Exp. Psychol. **34**(4), 859–873.

Bialystok, E., Craik, F., and Ryan, J. (**2006**). "Executive control in a modified anti-saccade task: Effects on aging and bilingualism," J. Exp. Psychol. **32**, 1341–1354.

Boersma, P. (**2001**). "PRAAT, a system for doing phonetics by computer," Glot International **5**(9/10), 341–345.

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-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**(4), 2339–2349.

Bradlow, A. R., Clopper, C., Smiljanic, R., and Walter, M. A. (**2010**). "A perceptual phonetic similarity space for languages: Evidence from five native language listener groups," Speech Commun. **52**, 930–942.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**(3), 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**(5), 2527–2538.

Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., and Bradlow, A. R. (**2009**). "Non-native speech perception in the presence of competing speech noise," Talk presented at the American Speech-Language-Hearing Association (ASHA), November 20, New Orleans, LA.

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., Van Engen, K. J., Dhar, S., and Bradlow, A. R. (**2010**). "The effects of clear speech as a masker," J. Speech, Lang. Hear. Res. **53**(6), 1–14.

Carhart, R., Tillman, W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**(3), 694–703.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., and La Heij, W. (**2008**). "How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms," J. Exp. Psychol. **34**(2), 302–312.

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**(1), 414–427.

Costa, A., and Santesteban, M. (**2004**). "Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners," J. Mem. Lang. **50**, 491–511.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Dixon, P. (**2008**). "Models of accuracy in repeated-measures design," J. Mem. Lang. **59**, 447–456.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**(5 Pt 1), 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**(5), 2246–2256.

Garcia 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.

IEEE Subcommittee (**1969**). "IEEE Subcommittee on Subjective Measurements IEEE Recommended Practices for Speech Quality Measurements," IEEE Transactions on Audio and Electroacoustics, **17**, 227–246.

Kidd G. Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach N. I., "Informational masking," in *Springer Handbook of Auditory Research 29: Auditory Perception of Sound Sources*, edited by W. Yost (Springer, New York, 2007), pp. 143–190.

Luk, G., de Sa, E., and Bialystok, E. (**2011**). "Is there a relation between onset age of bilingualism and cognitive control?" Bilingualism: Lang. Cognit. **14**(4), 588–595.

Mattys, S. L., Brooks, J., and Cooke, M. (**2009**). "Recognizing speech under a processing load: Dissociating energetic from informational factors," Cogn. Psychol. **59**, 203–243.

Mattys, S. L., Carroll, L. M., Li, C. K. W., and Chan, S. L. Y. (**2010**). "Effects of energetic and informational masking on speech segmentation by native and non-native speakers," Speech Commun. **11**, 887–899.

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

Meuter, R., and Allport, A. (**1999**). "Bilingual language switching in naming: Asymmetrical costs of language selection," J. Mem. Lang. **40**, 25–40.

Nábêlek, A. K., and Donahue. A. M., (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.

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

Pollack, I. (**1975**). "Auditory informational masking," J. Acoust. Soc. Am. **57**, S5.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (**1992**). "Temporary suppression of visual processing in an RSVP task: An attentional blink?" J. Exp. Psychol. **18**, 849–860.

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**, 1274–1277.

Rogers, C. L., Lister, J. J., Febor, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.

Rogers, R. D., and Monsell, S. (**1995**). "Costs of a predictable switch between simple cognitive tasks," J. Exp. Psychol. **124**(2), 207–231.

Russo, F., and Pichora-Fuller, M. K. (**2008**). "Tune in or tune out: Age-related differences in listening when speech is in the foreground and music is in the background," Ear Hear. **29**, 746–760.

Tun, P. A., O'Kane, G., and Wingfield, A. (**2002**). "Distraction by competing speech in young and older adult listeners," Psychol. Aging **17**(3), 453–467.

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

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.