Masking release due to linguistic and phonetic dissimilarity between the target and masker speech

Lauren Calandruccio[1,*], **Susanne Brouwer**[2], **Kristin J. Van Engen**[2,°], **Sumitrajit Dhar**[3], and **Ann R. Bradlow**[2]

[1]Department of Linguistics and Communication Disorders, Queens College of the City University of New York, Flushing, NY 11367

[2]Department of Linguistics, Northwestern University, Evanston, IL 60208

[3]Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208

Abstract

Purpose—To investigate masking release for speech maskers for linguistically and phonetically close (English and Dutch) and distant (English and Mandarin) language pairs.

Method—Twenty monolingual speakers of English with normal-audiometric thresholds participated. Data are reported for an English sentence recognition task in English, Dutch and Mandarin competing speech maskers (Experiment I) and noise maskers (Experiment II) that were matched either to the long-term-average-speech spectra or to the temporal modulations of the speech maskers from Experiment I.

Results—Results indicated that listener performance increased as the target-to-masker linguistic distance increased (English-in-English < English-in-Dutch < English-in-Mandarin).

Conclusions—Spectral differences between maskers can account for some, but not all, of the variation in performance between maskers; however, temporal differences did not seem to play a significant role.

Keywords

masking; native and non-native English speech perception

I. INTRODUCTION

Recognizing speech in the presence of competing speech can be difficult. Speech-recognition performance in noise can improve for listeners (i.e., they benefit from a masking release) when the relationship between the target and competing stimuli is manipulated (see Miller and Licklider, 1950; Festen and Plomp, 1990; Helfer and Freyman, 2008; Bernstein and Grant, 2009; and many others). Of particular interest for the present study is the suggestion from previous work that manipulations in the linguistic content of a masking speech signal can have a substantial influence on recognition of speech in the target signal. A masking release, or a decrease in overall masking, has been reported when the competing speech signal contained syntactically normal, but grammatically anomalous speech rather

Corresponding author information: Lauren Calandruccio, Department of Allied Health Sciences, The University of North Carolina at Chapel Hill. lauren_calandruccio@med.unc.edu.
[*]current affiliation Department of Allied Health Sciences, The University of North Carolina at Chapel Hill
[°]current affiliation Department of Psychology, Washington University in St. Louis

than meaningful linguistic content (Brouwer, Van Engen, Calandruccio, Dhar and Bradlow, 2012). In addition, several studies have reported a release from masking for first (L1) and second language (L2) speech perception when the target speech and masker speech were not spoken in the same language (e.g., Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio, Van Engen, Dhar, and Bradlow, 2010). Even non-natives attending to their L2 obtained a masking release when the competing speech was changed from their L2 to their L1 (i.e., they benefitted when the target and masker speech were not spoken in the same language, regardless of their proficiency in the two different competing languages; see Van Engen, 2010 and Brouwer et al., 2012).

It was hypothesized that since a release from masking has been observed for both speech maskers that are less meaningful as well as linguistically different from the target speech, a difference in the magnitude of masking release should be observed when the competing speech varies along a continuum in the degree of linguistic/phonetic similarity to the target speech. Testing this hypothesis will further our understanding of the contributions of linguistic/phonetic information to overall masking that could potentially improve signal-processing strategies within assistive listening devices for hearing-impaired listeners.

The goal of this research was to investigate masking release for foreign speech maskers that varied in the degree of linguistic/phonetic similarity to the target speech. Specifically, we were interested in comparing the magnitude of the masking release for linguistically and phonetically close (English and Dutch) and distant (English and Mandarin) language pairs. The three degrees of target-masker linguistic similarity included: (a) identical target-masker (English-in-English recognition), (b) linguistically close target-masker (English-in-Dutch recognition), and (c) linguistically distant target-masker (English-in-Mandarin recognition). We predicted that listeners would obtain a greater masking release when the competing language was more distant from the target speech than when it was close, because there should be greater differences in linguistic sound structure at the level of the phoneme inventories, syllable- and phrase-level phonetic structures, and rhythmic structure (and in turn, less overall masking). That is, we predicted that even when meaning was removed from the speech signal, the degree of similarities in such variables as rhythm class, phonemes, and syllable structures would be positively related to the extent of confusion between the target and masker signals. Data that supports this prediction will be presented. A follow-up investigation of the influence of spectral and temporal differences between the maskers will also be presented.

II. EXPERIMENT I: Linguistically and phonetically close and distant masker pairs

A. METHODS

Listeners—Twenty normal-hearing listeners (audiometric thresholds < 25 dB HL bilaterally at octave frequencies between 250 and 8000 Hz) participated in the experiment. All listeners were monolingual speakers of American English and included 13 females and 7 males ($M\ age$ = 21 years, SD = 2.4 years). Listeners were recruited from the student body at Northwestern University in Evanston, IL and were paid for their participation.

Stimuli—Target stimuli included sentences from the Bamford-Kowal-Bench (BKB) sentence lists (Bench, Kowal and Bamford, 1979; ® Cochlear Corporation) spoken by a native-English female speaker and recorded at Northwestern University. An example from the BKB sentences is, "The clown had a funny face", in which the keywords used for scoring are underlined.

The competing speech stimuli consisted of three different two-talker maskers, spoken in English, Dutch, and Mandarin. The two non-English masker languages differ from the target language, English, in various ways (see Table I). For example, Dutch and English are both from the West Germanic language family and have similar rhythm (both traditionally considered stress-timed) and phonotactics (wide range of permissible syllable structures). Mandarin is a Sino-Tibetan language; it has a much more restricted range of syllable structures (primarily CV syllables) compared to English and Dutch, and is a tonal language. During the experiment, subjects were also tested using a Croatian masker and a semantically anomalous English masker, but these results are not reported in this manuscript (see Calandruccio, Van Engen, et al., 2010 for a reported masking release for native-English speaking listeners listening to English in the presence of Croatian two-talker babble compared to English two-talker babble; see Brouwer et al. (2012) for results on masker effectiveness for meaningful and anomalous speech).

The Dutch sentences used during testing in Experiment I were direct translations (made by the second author who is a native-Dutch speaker) of the Nye and Gaitenby (1974) sentences that are syntactically correct but semantically anomalous. An example of these sentences is: The great car met the milk. An example of the same sentence translated into Dutch is: De geweldige auto ontmoette de melk. The Mandarin sentences, originally used in Van Engen and Bradlow (2007), are also syntactically correct, but semantically anomalous materials. The English masker consisted of syntactically correct, meaningful sentences spoken in English taken from the Harvard/IEEE sentence lists (IEEE, 1969). An example of a sentence from these lists is: Rice is often served in round bowls. It should be noted that though the English competing sentences were meaningful whereas the Dutch and Mandarin competing sentences were semantically anomalous, all listeners were monolingual speakers of English and had no knowledge of either Dutch or Mandarin. Brouwer et al. (2012) reported data for monolingual English listeners in the presence of meaningful and anomalous Dutch maskers. Results indicated no significant differences between the masker conditions; therefore, we would expect that since the listeners in the present study were all monolingual English speakers the fact that Dutch and Mandarin maskers were anomalous should not matter.

Six different female voices were used to create the three two-talker maskers (two native speakers each of English, Dutch, and Mandarin). The two-talker maskers were created by concatenating sentences spoken by each talker with no silent intervals between sentences. Though each of the two talkers spoke the same sentences in each language, the order of concatenation differed between the talkers in each masker condition. The sentences were equalized to the same root-mean-square (RMS) pressure level using Praat (Boersma and Weenink, 2012) prior to concatenation. The two strings of sentences were combined into a single audio file using Audacity®. The final audio files (one for each masker condition) were RMS equalized to the same overall pressure. Lastly, the ends of the audio files were digitally trimmed so that all three maskers were 34 seconds in length.

Instrumentation—The target and masker speech were mixed in real time using custom software created using MaxMSP (distributed by Cycling '74) running on an Apple Macintosh computer. Stimuli were passed to a MOTU 828 MkII input/output firewire device for digital-to-analog conversion (24 bit), passed through a Behringer Pro XL headphone amplifier and output to MB Quart 13.01HX drivers. Stimuli were then presented to the listeners via disposable foam insert earphones (13 mm) while seated in a comfortable chair within a double-walled sound-treated audiometric suite.

Experimental Testing—Listeners first participated in a pre-experiment with an easier signal-to-noise ratio (SNR) of −3 dB on the same day of testing. This experience allowed our listeners to be very comfortable with the speech-in-speech task and very familiar with

the target voice. Also, these initial 80 practice trials helped to alleviate learning effects within listeners' performance (see Felty, Buchwald, Pisoni, 2009).

Throughout testing, the level of the target speech remained fixed at 65 dB SPL, while the level of the competing (two-talker masker) speech was fixed at 70 dB SPL, resulting in a −5 dB SNR. The presentation order of the masker conditions (English, Dutch, Mandarin) was randomly varied across listeners and 16 sentences (1 BKB list; 50 keywords) were presented per masker condition.

Stimuli were presented binaurally. One target sentence was presented to the listener on each trial and a random portion of the appropriate two-talker masker was chosen and presented one second longer in duration compared to the target sentence (500 ms prior to the beginning of the target sentence, and 500 ms at the end of the target sentence). Listeners were asked to orthographically record what they heard on each trial. The written responses were scored as incorrect if the keyword was missing, incomplete, morphologically incorrect, or just wrong. Incorrect spelling of a word, however, was not considered incorrect.

B. RESULTS

The following statistical analyses are based on percent-correct data. The analysis was conducted to test whether English-sentence recognition differed among the three two-talker masker conditions. A mixed effects model with listener as a random variable was utilized (Baayen, Davidson, Bates, 2008). The fixed effect of masker was significant ($F = 36.04$, $p < .0001$). The least square means (LSM) for the three maskers were English = 38.9 ($SE = 3.29$), Dutch = 56.4 ($SE = 3.29$) and Mandarin = 72.4 ($SE = 3.29$). A post-hoc LSM Differences Tukey Honestly Significant Difference (HSD) test (Tukey, 1953) indicated a significant grouping difference between all three maskers. Data are illustrated in Figure 1 using boxplots. The length of the box indicates the interquartile range of performance scores, while the intermediate horizontal line indicates the median. The whiskers are calculated using the following two formulae: upper whisker = 3rd quartile + 1.5*(interquartile range), lower whisker = 1st quartile − 1.5*(interquartile range).

A post-hoc analysis was conducted to examine masking release relative to the most difficult condition (i.e. the English masker condition). Specifically, masking release was calculated by taking the within-participant difference in performance scores between (a) the Dutch and English masker conditions and, (b) the Mandarin and English masker conditions. A mixed effects regression model with subject as a random variable was conducted to test for a difference in masking release between Dutch-English and Mandarin-English. Results indicated a significant effect in masker language with respect to masking release ($p = .0099$). That is, there was a significantly larger masking release for the Mandarin-English condition, than the Dutch-English condition (see Figure 2). In addition, one-way t-tests also indicated that the masking release observed for both Dutch and Mandarin were significantly different than zero ($t_{(19)} = 3.59$, $p = .0019$ and $t_{(19)} = 9.99$, $p < .0001$, respectively).

C. DISCUSSION

Data from monolingual English speakers indicate that when listening to English sentences in competing speech, a competing English masker is most effective, followed by Dutch, and further followed by Mandarin. These data support the original hypothesis that masker effectiveness for a target signal decreases as the competing speech becomes more distant phonetically from the target speech compared to competing speech that is (more) similar to the target language. These data suggest that similar phonemes, phonotactics, and other phonetic or phonological structure similarities between a target and a masker speech signal can increase overall masking. However, it must be considered that the different voices used

to create the two-talker maskers had different spectral and temporal properties. A close examination of the long-term average speech spectra (LTASS) between the three maskers can be observed in Figure 3. The Mandarin masker has noticeably less energy above 5000 Hz than the English and Dutch maskers. Therefore, it is possible that differences other than those that are linguistically driven between the maskers might have contributed to the significant results observed in Experiment I. The purpose of Experiment II was to attempt to isolate some of these potential spectral-temporal signal-related features across the three two-talker maskers.

III. EXPERIMENT II: Spectrally matched steady-state and temporally modulated white-noise maskers

In an attempt to examine spectral and temporal differences between the masker conditions that could potentially be impacting the results of Experiment I, a second experiment was conducted using noise (rather than speech) maskers. Two different sets of noise maskers were created. The first set of noise maskers were spectrally matched to the three two-talker maskers (English, Dutch, and Mandarin) used in Experiment I. This manipulation removed temporal differences between the three maskers, while preserving the long-term spectral content of the original maskers. The second set of noise maskers included three white-noise maskers temporally modulated to match the low-frequency modulations of the three two-talker maskers used in Experiment I. Thus, this manipulation removed all spectral differences between the three maskers, but preserved the low-frequency temporal modulations of the original two-talker maskers.

A. METHODS

Listeners—Twelve additional native-English speaking normal-hearing (audiometric thresholds < 25 dB HL bilaterally at octave frequencies between 250 and 8000 Hz) listeners (11 females and 1 male) participated in Experiment II (M age = 23 years, SD = 2.8 years). Listeners were recruited from the student body at Queens College at the City University of New York. Participants signed an informed consent form approved by the IRB at Queens College and were paid for their participation.

Stimuli—Target stimuli were taken from the same BKB sentences used in Experiment I. The three steady-state (SS) noise maskers were spectrally matched to the three two-talker maskers and were generated in MATLAB by passing a Gaussian white noise through an FIR filter with 2048 points and a magnitude response equal to each individual LTASS of the three two-talker maskers. The temporally modulated (TM) white-noise maskers were computed using MATLAB. A full-wave rectification Hilbert transform was applied to the three speech maskers used in Experiment I. Stimuli were then low-pass filtered using a rectangular filter that utilized a sampling rate of 22.1 kHz and a cutoff frequency of 50 Hz (see Davidson, Gilkey, Colburn, and Carney, 2006). A Gaussian white noise, also generated in MATLAB, was then multiplied by the different envelopes to create three TM noise maskers (one for each of the three original speech maskers used in Experiment I). The SS spectrally matched noise maskers and the TM white-noise maskers were then RMS equalized to the same pressure level as the target sentences using Praat (for similar methods see Calandruccio, Dhar, and Bradlow, 2010).

Procedure—The target and masker were mixed in real time using custom software created with MaxMSP running on an Apple Macintosh computer. Stimuli were passed to a MOTU UltraLite input/output firewire device for digital-to-analog conversion (24 bit), to an Art HeadAmp6Pro headphone amplifier and output to Etymotic ER1 insert earphones. Stimuli

were then presented to the listeners via disposable foam insert earphones (13 mm) while seated in a comfortable chair within a double-walled sound-treated audiometric room.

Experimental Testing—All listeners initially came in for testing using the SS spectrally matched noise maskers. Two months later, the same listeners returned for testing using the TM white-noise maskers. Procedures used in Experiment II were similar to those used in Experiment I. All maskers were presented at a fixed SNR of −5 dB (the same SNR used in Experiment I). For both types of noise maskers, one practice BKB list was used to familiarize the listener with the task. The presentation order of the masker conditions was randomly varied across listeners, and 16 sentences (1 BKB list; 50 keywords) were presented per masker condition.

B. RESULTS

A mixed effects model with listener as a random variable was utilized for both masker types (SS and TM). These analyses were conducted to test whether English-sentence recognition differed when listening between the SS and TM noise maskers (based on the three two-talker masker conditions). The fixed effect of masker was significant for the spectrally matched SS masker ($F = 13.63$, $p < .0001$). The LSM for the three maskers were English = 61.61 ($SE = 3.28$), Dutch = 69.83 ($SE = 3.14$) and Mandarin = 83.08 ($SE = 3.14$). A post-hoc LSM Differences Tukey HSD test indicated the Mandarin masker was significantly less effective compared to the English and Dutch maskers; no significant grouping difference was observed between the English and Dutch maskers. For the TM masker there was not a significant fixed effect of masker [$F = 13.63$, $p = .3433$ (see Figure 4)].

As for Experiment I post-hoc analysis was conducted to examine masking release with respect to the most difficult condition (i.e. the respective English SS and English TM masker conditions). Specifically, masking release was calculated by taking the difference in performance scores between the Dutch and English SS, Mandarin and English SS, Dutch and English TM and Mandarin and English TM masker conditions. A mixed effects regression model with subject as a random variable was conducted to test masking release between the masker conditions. Specifically, we examined making release for Dutch-English and Mandarin-English for both the SS and TM masker types and the interaction between masker language and masker type. The main effect of masker language was not significant ($p = .1881$), however, the main effect of masker type was significant ($p < .0001$). The interaction between language and masker type was significant ($p = .0228$). Post-hoc Tukey HSD testing indicated significant grouping differences for the Mandarin-English SS masking release compared to the masking release for the two TM masker conditions. There was not a significant grouping difference for the masking release observed between the Dutch-English TM maskers and Mandarin-English TM maskers. Nor was there a significant difference between the masking release observed for between the Dutch-English SS and Mandarin-English SS maskers. Figure 5 illustrates the masking release that was observed between the English and the two foreign language SS and TM maskers. For the SS maskers additional one-way t-tests were conducted and indicated that the masking release observed for the Dutch-English condition was not significantly different from zero ($t_{(11)} = 1.39$, $p = .195$), while the Mandarin-English masking release was significant ($t_{(11)} = 6.13$, $p < .0001$). A key comparison to make is that in Figure 2 (above) there was a significant masking release for the Dutch-English two-talker masker. Therefore, these data, taken in combination with those reported in Figure 2 support the conclusion that some portion of the masking differences in the two-talker condition, particularly for the Dutch two-talker masker, cannot be traced to spectral or temporal differences amongst the maskers.

C. DISCUSSION

The spectral energy within the Mandarin masker was less effective in masking the English target speech than the other two maskers (English and Dutch). These results indicate that, at a minimum, a portion of the Mandarin masker ineffectiveness observed in Experiment I was due to energetic masking differences between maskers (and not solely linguistic and or phonetic distance). Differences in temporal modulations between the three speech maskers used in Experiment I alone were not large enough to cause significant differences in recognition performance.

The results from Experiment I cannot be fully explained based on linguistic and phonetic distance between the target (English) and the masker (English, Dutch or Mandarin) since the energetic masking contributions between the Dutch and English maskers and the Mandarin masker were not equal. However, the data from Experiment II also indicate that the difference in masker effectiveness that was observed between the English and Dutch speech maskers in Experiment I cannot be accounted for solely by energetic masking differences.

III. GENERAL DISCUSSION

Decreasing the similarity between the target and masker speech (i.e., English targets and Mandarin masker) allowed for a greater release from masking relative to a more phonetically similar target and masker combination (i.e., English targets and either English or Dutch maskers). In other words, there was a gradient improvement in performance for English-in-English to English-in-Dutch to English-in-Mandarin listening conditions. Though the data from Experiment I suggest that linguistic and phonetic dissimilarity had a linear relationship with masker effectiveness, the data from Experiment II strongly suggest that a portion of the ineffectiveness observed for the Mandarin masker was due to reduced spectral overlap with the target speech.

In our study, two stress-timed languages were used: English and Dutch. Mandarin is a syllable-timed language (Lin and Wang, 2005). Reel (2009) reported that speech maskers were less effective when the masking speech was dissimilar in terms of rhythmic structure relative to the target speech, especially when the rhythm class of the masker was unknown to the listener. Therefore, the differences in rhythm between the maskers relative to the target English speech may also have contributed to the Mandarin masker being the least effective speech masker. It should be noted that rhythm class may not be synonymous with differences in low-frequency envelope modulations because rhythm class is a linguistic rather than acoustic classification that relates to the internal shape of syllables (including factors such the presence of absence in the language of a phonological process such as vowel reduction in unstressed syllables and consonant cluster permissibility) which do not directly and straightforwardly relate to envelope modulations in the speech signal. That is, target and masker speech with different rhythms may be easier to segregate than target and masker speech with similar rhythm. However, further research is needed to determine if the differences in rhythm class with respect to masker effectiveness has to do with differences in low-frequency envelope modulations (i.e., greater or more frequent "dips") or sound-source segregation.

The data reported in Experiments I and II illustrate the importance for researchers to be consistent and thorough in reporting temporal and spectral properties of the signals used within speech masking experiments. To make valid interpretations of non-energetic masking effects caused by linguistic maskers, energetic contributions must be understood first.

These data are preliminary in confirming that masking release progressively increases with progressive changes in target and masker linguistic similarity. It is possible that the

differences observed in these experiments are speaker specific and not language specific. As we move forward we must find ways to control for spectro-temporal differences between different speech maskers without simultaneously eliminating critical linguistic-phonetic information. Possible ways to minimize energetic differences include having the same talkers create multiple masker conditions (Freyman, Balakrishnan, and Helfer, 2001) or normalizing LTASS between the masker conditions (Brouwer et al., 2012). Using three competing talkers as opposed to two may also help to reduce temporal differences that may exist between masker conditions that can improve or degrade masker effectiveness (Calandruccio, Dhar, et al., 2010) while still allowing to probe non-energetic masking effects (Freyman, Balakrishnan, Helfer, 2004; Simpson and Cooke, 2005). If the specific properties within competing signals that allow for greater masking release could be identified, there is potential that those properties could eventually be incorporated into signal processing strategies. In turn, those strategies could help improve speech recognition when listening in noise for those with hearing loss.

Acknowledgments

Acronyms and/or abbreviations

SNR	(signal-to-noise ratio)
BKB	(Bamford-Kowal-Bench)
IRB	(Insitutional Review Board)
SPL	(Sound Pressure Level)
M	(mean)
SD	(standard deviation)
SE	(standard error)
long-term-average speech spectra	(LTASS)
L1	(first language)
L2	(second language)
steady state	(SS)
temporally modulated	(TM)
least square means	(LSM)
root-mean-square	(RMS)
honestly significant difference	(HSD)

References

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang . 2008; 59:390–412.

Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br. J. Audiol . 1979; 13(3):108–12. [PubMed: 486816]

Bernstein JG, Grant KW. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. J. Acous. Soc. Am. 2009; 125(5):3358–72.

Boersma, P.; Weenink, D. Praat: doing phonetics by computer [Computer program]. 2012. Version 5.3.15, retrieved 19 May 2012 from http://www.praat.org/

Booij, G. The phonology of Dutch. Oxford University Press; Oxford: 1995.

Brouwer S, Van Engen KJ, Calandruccio L, Dhar S, Bradlow AR. Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. J. Acous. Soc. Am. 2012; 131(2):1449–64.

Calandruccio L, Dhar S, Bradlow AR. Speech-on-speech masking with variable access to the linguistic content of the masker speech. J. Acous. Soc. Am. 2010; 128(2):860–9.

Calandruccio L, Van Engen K, Dhar S, Bradlow AR. The effectiveness of clear speech as a masker. J. Sp. Lang. Hear. Res. 2010; 53(6):1458–71.

Davidson SA, Gilkey RH, Colburn HS, Carney LH. Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results. J. Acous. Soc. Am. 2006; 119(4):2258–75.

Dryer, MS.; Haspelmath, M., editors. The World Atlas of Language Structures Online. Max Planck Digital Library; Munich: 2011. Available online at http://wals.info/

Felty RA, Buchwald A, Pisoni DB. Adaptation to frozen babble in spoken word recognition. J. Acoust. Soc. Am. 2009; 125(3):EL93–7. [PubMed: 19275281]

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J. Acoust. Soc. Am. 1990; 88(4):1725–1736. [PubMed: 2262629]

Freyman RL, Balakrishnan U, Helfer KS. Spatial release from informational masking in speech recognition. J. Acous. Soc. Am. 2001; 109(5):2112–22.

Freyman RL, Balakrishnan U, Helfer KS. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Am. 2004; 115(5):2246–2256. [PubMed: 15139635]

Garcia Lecumberri ML, Cooke M. Effect of masker type on native and non-native consonant perception in noise. J. Acous. Soc. Am. 2006; 119(4):2445–54.

Helfer KS, Freyman RL. Aging and speech-on-speech masking. Ear Hear. 2008; 29(1):87–98. [PubMed: 18091104]

IEEE Subcommittee. IEEE Subcommittee on Subjective Measurements IEEE Recommended Practices for Speech Quality Measurements. IEEE Transactions on Audio and Electroacoustics. 1969; 17:227–246.

Li, C.; Thompson, SA. Mandarin Chinese: A functional reference grammar. University of California Press; Berkeley and Los Angeles, CA: 1989.

Lin H, Wang Q. Mandarin rhythm: an acoustic study. J. Chinese Lang. and Computing. 2005; 17:127–140.

Miller GA, Licklider JCR. The intelligibility of interrupted speech. J. Acous. Soc. Am. 1950; 22:167–173.

Nye, PW.; Gaitenby, JH. Status Report on Sp. Res. Haskins Laboratory; 1974. The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. SR-37/38

Reel, L. PhD Thesis. Texas Tech University Health Sciences Center; Lubbock, TX, USA: 2009. Selective auditory attention in adults: Effects of rhythmic structure of the competing language.

Simpson S, Cooke MP. Consonant identification in N-talker babble is a nonmonotonic function of N. J. Acous. Soc. Am. 2005; 118:2775–2778.

Tukey JW. Some selected quick and easy methods of statistical analysis. Trans. N. Y. Acad. Sci. 1953; 16(2):88–97. [PubMed: 13169230]

Van Engen KJ. Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. Speech Commun. 2010; 52(11-12):943–53. [PubMed: 21179561]

Van Engen KJ, Bradlow AR. Sentence recognition in native- and foreign-language multi-talker background noise. J. Acous. Soc. Am. 2007; 121(1):519–26.
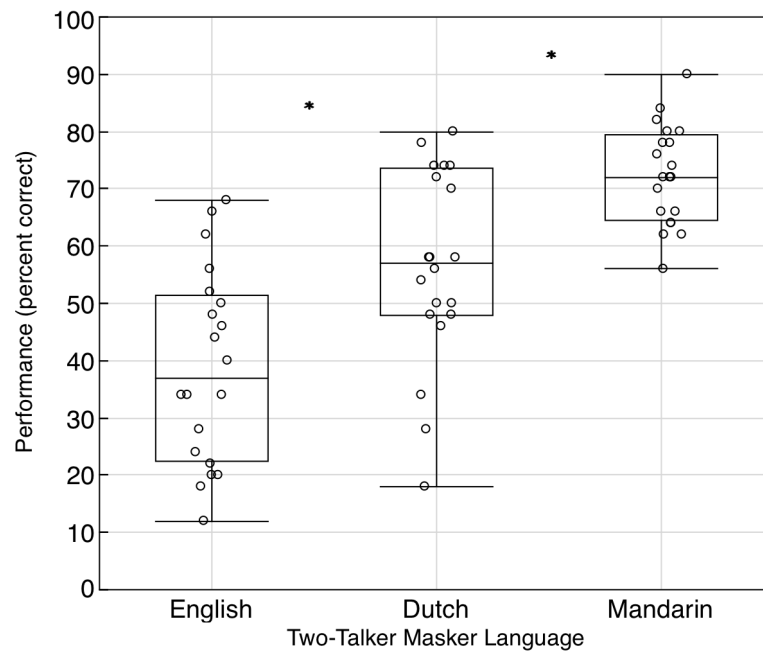
Figure 1.
Sentence recognition performance (percent correct) in the presence of three two-talker maskers spoken in English, Dutch and Mandarin. Boxplots for each linguistic masker are shown. The length of the box indicates the interquartile range of performance scores, while the intermediate horizontal line indicates the median. The whiskers are calculated using the following two formulae: upper whisker = 3rd quartile + 1.5*(interquartile range), lower whisker = 1st quartile − 1.5*(interquartile range). Individual data points are also indicated for all 20 listeners within each boxplot.
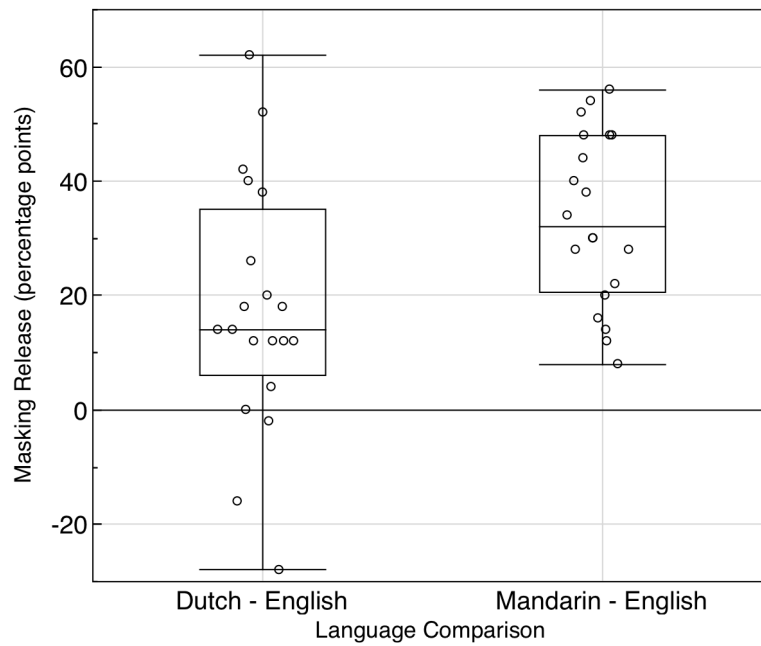
Figure 2.
Masking release for data reported in Experiment I. Masking release was calculated by subtracting each subject's sentence recognition performance in the presence of the foreign language masker minus their performance in the English masker (i.e., Dutch minus English, and Mandarin minus English). The magnitude of the masking release was significantly different between the two foreign languages. Specifically, Mandarin allowed for a significantly greater masking release than Dutch. The masking release for both languages was significant.
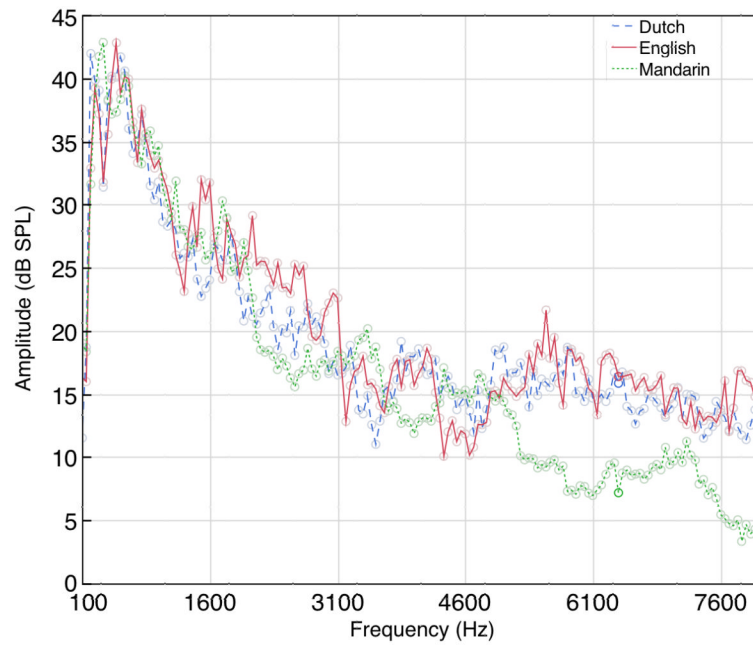
Figure 3.
Long-term average speech spectra for the three linguistic maskers used in Experiment I. The Mandarin masker has noticeably less energy above 5000 Hz than the English and Dutch maskers.
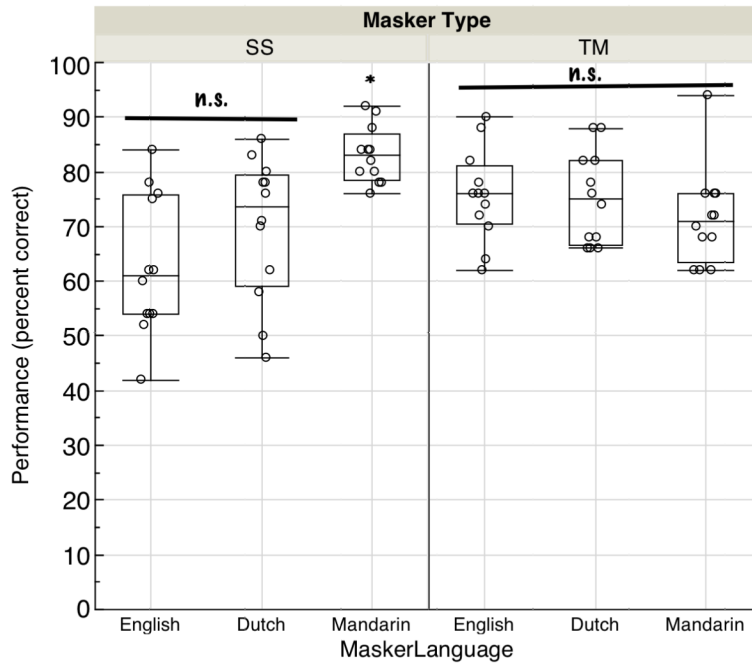
Figure 4.
Sentence recognition performance (percent correct) in the presence of steady state (SS) and temporally modulated (TM) noise maskers derived based on the original three two-talker English, Dutch and Mandarin maskers used in Experiment I. Boxplots are shown for each masker condition. Similarly to Figure 1, the length of the box indicates the interquartile range of performance scores, while the intermediate horizontal line indicates the median. The whiskers are calculated using the following two formulae: upper whisker = 3rd quartile + 1.5*(interquartile range), lower whisker = 1st quartile − 1.5*(interquartile range). Individual data points are also indicated for all 12 listeners within each boxplot.
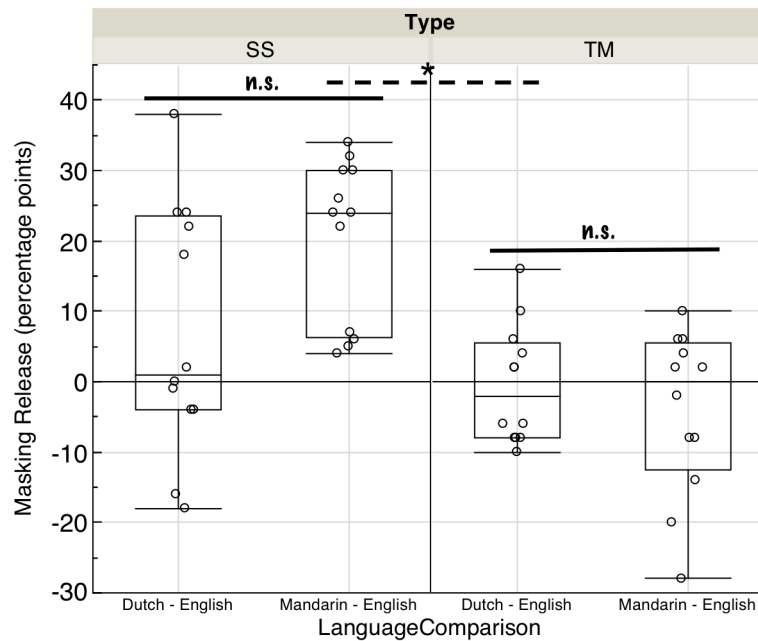
Figure 5.
Masking release for data reported in Experiment II. Masking release was calculated by subtracting each subject's sentence recognition performance in the presence of the foreign language SS and TM masker minus their performance in the English SS and TM masker, respectively (i.e., Dutch SS minus English SS and Mandarin SS minus English SS; Dutch TM minus English TM and Mandarin TM minus English TM). The masking release observed for the Mandarin SS – English SS masker was significantly greater than the masking release observed for the TM masker conditions. However, there was not a significant difference in masking release between the two SS masker comparisons, nor the two TM masker comparisons (indicated by 'n.s.'). The significant interaction between language and masker type is indicated by an '*'.

Table I

Languages used for the masker conditions

Language of the masker	# of Vocalic Phonemes	# of Consonantal Phonemes	Linguistic Family	Syllable Structure	Lexical Tones	Rhythm Class
English[1]	14	24	Indo-European (West Germanic)	$(C)^3 V(C)^4$	No	Stress-timed
Dutch[2]	13	26	Indo-European (West Germanic)	$(C)^3 V(C)^4$	No	Stress-timed
Mandarin[3]	35	28	Sino-Tibetan	$(C)V(C)$	Yes	Syllable-timed

[1] Dryer and Haspelmath, 2011

[2] Booij, 1995

[3] Li and Thompson, 1989