Sentence recognition in native- and foreign-language multi-talker background noise[a)]

Kristin J. Van Engen[b)] and Ann R. Bradlow
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 18 April 2006; revised 25 October 2006; accepted 25 October 2006)

Studies of speech perception in various types of background noise have shown that noise with linguistic content affects listeners differently than nonlinguistic noise [e.g., Simpson, S. A., and Cooke, M. (2005). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778; Sperry, J. L., Wiley, T. L., and Chial, M. R. (1997). "Word recognition performance in various background competitors," J. Am. Acad. Audiol. **8**, 71–80] but few studies of multi-talker babble have employed background babble in languages other than the target speech language. To determine whether the adverse effect of background speech is due to the linguistic content or to the acoustic characteristics of the speech masker, this study assessed speech-in-noise recognition when the language of the background noise was either the same or different from the language of the target speech. Replicating previous findings, results showed poorer English sentence recognition by native English listeners in six-talker babble than in two-talker babble, regardless of the language of the babble. In addition, our results showed that in two-talker babble, native English listeners were more adversely affected by English babble than by Mandarin Chinese babble. These findings demonstrate informational masking on sentence-in-noise recognition in the form of "linguistic interference." Whether this interference is at the lexical, sublexical, and/or prosodic levels of linguistic structure and whether it is modulated by the phonetic similarity between the target and noise languages remains to be determined. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400666]

PACS number(s): 43.71.Es, 43.71.Hw, 43.72.Dv [PEI] Pages: 519–526

I. INTRODUCTION

The substantial literature on speech-in-noise perception has been successful in revealing the relative resistance of various speech signal features to degradation from noise, as well as in assessing the relative abilities of various listener populations to recover from the detrimental effects of background noise. A particularly noteworthy finding of several recent linguistic and audiological studies is that the presence of background noise can force a "re-ranking" of acoustic cues to linguistic categories such that "secondary" cues in quiet become the only available, and hence "primary," cues in noise (Parikh and Loizou, 2005; Jiang *et al.*, 2006). Mattys *et al.* (2005) also provide evidence that listeners assign different weights to various cues for word segmentation when the speech signal is fully available versus degraded by noise. Similarly, the presence of noise can "re-rank" listener groups such that groups that perform equivalently in quiet may perform differently in noise (Nábělek and Donohue, 1984; Takata and Nábělek, 1990; Mayo *et al.*, 1997; Van Wijngaarden *et al.*, 2002; but see Cutler *et al.*, 2004 for comparable effects of noise on native and non-native listener phoneme identification). Furthermore, noteworthy discrepancies between quiet and noisy test conditions have been observed for intelligibility of native- versus foreign-accented speech (Rogers *et al.*, 2004). These findings suggest that listeners process speech signals differently when they are embedded in noise as opposed to in quiet, and that a comprehensive understanding of speech perception requires studies of speech perception under various noise conditions.

Accordingly, the present study investigated English sentence perception in the presence of multi-talker babble with varying numbers of talkers in the babble, varying signal-to-noise ratios (SNRs), and varying languages in the background noise. Since speech perception in noise is likely to be affected by a combination of lower-level (peripheral, energetic) masking and higher-level (central cognitive, linguistic, informational) masking, particularly in the case of background speech babble noise, there is likely to be a range of noise characteristics (some SNRs, types of noise) where the linguistic content of the noise has a direct influence on the recognition of target speech. If we can identify this range of noise characteristics, then we can begin to isolate the various linguistic features (fine-grained acoustic phonetic segment-level to lexical and higher-level prosodic) that are involved in speech-in-speech perception. Our overall interest is in developing a deeper understanding of the linguistic factors involved in speech-in-noise recognition.

A key strategy for investigating the effects of noise on speech processing and for ultimately developing a principled account of these effects is to compare different types of noise, which vary with respect to the kind and degree of interference they impose on speech signals. To this end, lin-

[a)]A version of this study was presented at the 150th meeting of the Acoustical Society of America, Minneapolis, MN, October 2005 and at the Mid-Continental Workshop on Phonology 11, University of Michigan, November 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: k-van@northwestern.edu

guistic and audiological studies have employed a wide variety of noise types, including single-talker maskers, multi-talker babble with various numbers of talkers, speech-shaped noise, and white noise. In general, these studies have shown that, regardless of the type of noise, performance on speech recognition tasks decreases as the level of the noise increases relative to the level of the target speech. With respect to speech noise in particular, they have shown that greater similarity between masker and target voices in terms of characteristics such as vocal tract size and fundamental frequency decreases intelligibility (Brungart *et al.*, 2001). Target intelligibility also generally decreases as additional voices are added to multi-talker babble (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005; Simpson and Cooke, 2005).[1] As for comparisons across speech and nonspeech noise, Simpson and Cooke (2005) found lower speech intelligibility scores in natural babble than in babble-modulated noise when there were more than two talkers in the noise (see Sperry *et al.*, 1997 for a similar result). This difference in the effects of natural babble and babble-modulated noise suggests that linguistic interference plays a role in the effects of natural multi-talker babble on target speech perception.

Although studies such as Simpson and Cooke (2005) have provided evidence for particularly linguistic effects of multi-talker noise, there has been little investigation of the factors involved in such effects. In multi-talker babble studies, for example, the language spoken in the babble has typically matched the language spoken in the target. As a consequence, we have limited information about the precise linguistic features—phonemes, words, prosodic characteristics—that are most responsible for the greater masking effects of speech noise than nonspeech noise.

Two recent studies, however, have used multiple noise languages in order to examine other aspects of noise and perception (Rhebergen *et al.*, 2005; Garcia Lecumberri and Cooke, 2006). Rhebergen *et al.* (2005) used two noise languages to examine the effects of time-reversing interfering speech. In general, target speech intelligibility is known to be better in time-reversed interfering speech than in forward interfering speech—an effect attributed to the removal of any interfering informational content in the noise. However, reversing speech also results in increased forward masking, which increases the energetic masking imposed by the noise on the target speech. Rhebergen *et al.* (2005) assessed the relative effects of these two opposing factors (reduced informational masking but increased energetic masking in time-reversed speech) by comparing the effects of forward and reversed Dutch babble with forward and reversed Swedish babble on the recognition of Dutch speech for Dutch listeners. A comparison of the Dutch-in-Dutch noise versus Dutch-in-Swedish noise (without time reversal) showed better speech reception thresholds in the Swedish noise condition.

Garcia Lecumberri and Cooke (2006) used two noise languages in a study of native and non-native listeners' perception of English consonants in noise. Their primary conclusion, based on a comparison of a variety of noise types, was that non-native listeners were more adversely affected than native listeners by both energetic and informational masking. A secondary finding, of direct relevance to the present study, was that the native English speakers performed slightly better in Spanish noise than in English noise. The present study expands on this finding with a more direct and systematic study of the effects of two noise languages on native listeners. Furthermore, by examining sentence intelligibility rather than consonant identification, this study involves more levels of linguistic knowledge and more closely represents real-world listening situations in which listeners must extract meaningful messages from noisy environments.

TABLE I. Aspects of native language versus foreign language two- and six-talker babble and their predicted effects on target speech intelligibility. "No" represents a feature of the noise that is expected to hinder target speech intelligibility. "Linguistic differentiation" refers broadly to differences in phonetic, phonological, lexical, and prosodic characteristics of the languages. It should be noted that the babble in this study was constructed from semantically anomalous sentences, so word transition probabilities and/or sentential semantics are not taken into account.

	Same language two-talkers	Different language two-talkers	Same language six-talkers	Different language six-talkers
Temporal gaps	Yes	Yes	No	No
Linguistic differentiation	No	Yes	No(?)	Yes

II. METHOD

This study compares the intelligibility of native-accented English sentences for native English listeners in the presence of English two- and six-talker babble versus Mandarin two- and six-talker babble at SNRs of +5, 0 and −5 dB.[2] Mandarin is particularly well suited to this investigation because it differs significantly from English with respect to several levels of linguistic structure—phoneme inventory, syllable structure, rhythmic properties, and prosodic properties. By comparing languages that differ dramatically, such as these, the chance of observing differential speech noise effects is maximized.

Table I lists aspects of two- and six-talker babble in the target speech language and in a different language that can be expected to affect target speech intelligibility for native speakers of the target language: (a) the amount/duration of temporal gaps in the noise and (b) the amount of linguistic differentiation between the target and the noise with respect to phonetic, phonological, lexical, and prosodic characteristics. Based on previous multi-talker babble findings, we expect that six-talker babble will be a more effective masker than two-talker babble.[3] With respect to linguistic characteristics, it is hypothesized that the degree to which linguistic information in the noise—individual phonemes, phonotactics, prosody, lexical items—matches the linguistic information in the target will correlate with the amount of interference caused by that noise type on the signal. For this reason, same-language noise should be more detrimental to target speech intelligibility than different-language noise.[4]

Listeners were presented with target sentences in English mixed with multi-talker babble in either English or Mandarin and were asked to write down what they heard.

TABLE II. Experimental design: conditions, block types, and block ordering.

	No. of Talkers in Noise	Block 1 Mandarin	Block 2 English	Block 3 Mandarin	Block 4 English
Condition 1	6	SNR: +5	SNR: +5	SNR: 0	SNR: 0
Condition 2	6	SNR: 0	SNR: 0	SNR: −5	SNR: −5
Condition 3	2	SNR: +5	SNR: +5	SNR: 0	SNR: 0
Condition 4	2	SNR: 0	SNR: 0	SNR: −5	SNR: −5

Four independent groups of listeners participated in four different experimental conditions, each one containing four blocks of trials, as shown in Table II. These conditions allowed for direct comparison of the effect of two versus six-talker babble (between subjects), the effect of English versus Mandarin babble (within subjects), and the effect of different signal-to-noise ratios (within subjects).

A. Participants

1. Speakers

Six monolingual native speakers of general American English (three males and three females between the ages of 28 and 48 years) and six native speakers of Mandarin Chinese (three males and three females between the ages of 24 and 37) provided recordings in their native languages to be used for the English and Mandarin noise tracks (described below). The English speakers were graduate students and postdoctoral researchers in the Northwestern University Linguistics Department, recorded for a previous experiment (Smiljanic and Bradlow, 2005). The Mandarin speakers were graduate students and family members of graduate students at Northwestern University. A different adult female speaker of general American English produced the target sentences.

2. Listeners

Seventy-three undergraduate participants were recruited from the Northwestern University Linguistics Department subject pool and received course credit for their participation in the study. Seven participants were omitted from the final analysis—two reported a hearing loss, two were non-native speakers of English, two were English-Mandarin bilinguals, and one was omitted due to computer error during data collection. The remaining 66 participants were native English speakers between the ages of 18 and 23, 18 of which were bilingual speakers of English and a language other than Mandarin. All reported having normal speech and hearing. The distribution of subjects across conditions was as follows: condition 1: $n=16$; condition 2: $n=17$; condition 3: $n=17$; condition 4: $n=16$.

B. Stimuli

1. Generating multi-talker babble

For the "noise" sentences, each speaker produced a set of 20 semantically anomalous sentences in either English (e.g., *Your tedious beacon lifted our cab; My puppy may stress their fundamental gallon*) or Mandarin. These English sentences were developed for unrelated research (Smiljanic and Bradlow, 2005), and were used in this study to eliminate the possibility that participants might extract an entire meaningful sentence from a speaker other than the target. The Mandarin sentences were direct translations of the English sentences (translated by one native Mandarin speaker and checked by another).

Participants were instructed to speak in a natural, conversational style, and to repeat any sentences in which they produced disfluencies. Recordings took place in a sound-attenuated booth in the phonetics laboratory of the Department of Linguistics at Northwestern University. Participants read the sentences from index cards and spoke into a microphone, recording directly to disk using an Apogee PSX-11 analog/digital and digital/analog converter. Recordings were digitized at a sampling rate of 16 kHz with 24 bit accuracy. Sentences were then separated into individual files and equated for rms amplitude so that they would all contribute equally to the babble.

English and Mandarin six-talker babble was created from these recordings as follows: for each talker, two sentences (a different pair of sentences for each talker) were concatenated to ensure the duration of the noise tracks would exceed the durations of all target sentences. A multiple of 100 ms of silence was added to each talker's file (0–500 ms) in order to stagger the talkers once they were mixed together. All six talkers were then mixed, and the initial 500 ms of the mixed file was removed to eliminate noise that did not contain all six talkers. The first 100 ms of the completed noise file was faded in, and the final noise file was leveled in rms amplitude to produce SNRs of +5, 0, and −5 dB when mixed with the stimulus sentences. The stimulus sentences were each leveled to the same rms amplitude (60 dB), and the relevant SNRs were produced by leveling each noise file relative to the level at which the sentence files had been leveled (55, 60, 65 dB).

For two-talker babble, two female voices were used for both English and Mandarin. This was done primarily to match the gender of the target speaker and thus eliminate the variable of gender differences in speech-in-speech intelligibility (see Brungart *et al.*, 2001). Furthermore, it was hoped that using the same gender for the two talkers would lead to better perceptual fusion of the pair so that it would be treated by listeners as two-talker babble. Four different two-talker noise tracks were generated for each language. Again, two sentences by each speaker were concatenated to ensure adequate duration of the noise file (four sentence pairs total per speaker). For one of the two speakers, 500 ms of silence was added to the beginning of the files. The two talkers were mixed as above and the first 500 ms were removed. Finally, the first 100 ms were faded in and the completed noise tracks were each leveled to the three rms amplitudes necessary to produce SNRs of −5, 5, and 0 dB when mixed with the leveled target sentences.

2. Target sentences

Target sentences for the present study were taken from a set of recordings originally made for an unrelated study (Bent and Bradlow, 2003). The sentences were taken from

the Revised Bamford-Kowal-Bench Standard Sentence Test, lists 7–10. Each list contains 16 simple, meaningful sentences (e.g., *The children dropped the bag; five men are working*) and 50 keywords (three or four per sentence) for a total of 64 sentences and 200 keywords. Lists 7, 8, 9, and 10 were selected based on their equivalent intelligibility scores for normal children as reported in Bamford and Wilson (1979). For additional details, see Bent and Bradlow (2003).

These recordings were mixed with the noise files, a 400 ms silent leader was inserted, followed by 500 ms of the noise alone, then the target signal mixed with the noise, and finally 500 ms of noise at the end of each trial. Each target sentence was mixed with each type of noise file, yielding 12 sets of target stimuli: two noise languages (English, Mandarin) X 2 talker numbers (two talker, six talker) X 3 SNRs (+5, 0, −5 dB). For six-talker noise, the same noise file was mixed with every target sentence at each SNR. For the two-talker noise, each of the four different tracks was used for 25% of the target sentences in each language and at each SNR.

C. Procedure

Listeners were seated in a sound-attenuated booth facing a computer monitor. Stimuli were presented diotically over headphones (Sennheiser HD 580) at a comfortable level. Participants were presented with a total of 68 trials—four practice sentences followed by four experimental blocks of 16 sentences each. They were instructed that they would be listening to sentences mixed with noise, and were asked to write down what they heard. They were asked to guess if they were unsure, and also to report individual words if that was all they could identify. The task was self-paced; participants pressed the space bar on the keyboard to advance from trial to trial. Participants could listen to each sentence no more than once. After the practice block, the experimenter verified that the equipment was functioning properly and checked the readability of the participant's handwriting.

Practice items (two sentences in English noise and two in Mandarin noise) were presented at the same SNR as block 1 (+5 dB for conditions 1 and 3, 0 dB for conditions 2 and 4). The sentences in block 1 were mixed with Mandarin noise; in block 2 the sentences were presented at the same SNR as block 1 but with English noise; in block 3 the sentences were presented at the more difficult SNR with Mandarin noise; and in the final block the sentences were presented at the more difficult SNR but with English noise (as shown in Table II).

English noise at the difficult SNR was predicted to be the most difficult block in all conditions due to the higher noise level and the greater linguistic overlap between the noise and the target. Therefore, this block was presented last, giving participants maximal opportunity to adjust to the task and to the target talker, thereby "stacking the cards" against our predicted result of better English sentence recognition with Mandarin noise than with English noise. Because of the possibility that some of the target sentences may be more or less easy to perceive than others, the four target sentence lists were counterbalanced across the four possible orderings. The ordering of blocks with respect to the type and level of noise was consistent for all participants (as described above).

D. Data analysis

Perception scores were determined by a strict keyword-correct count. Each set of 16 sentences contained 50 keywords, and listeners received credit for each keyword transcribed perfectly. Words with added or deleted morphemes were considered incorrect, but obvious spelling errors or homophones were counted as correct. Raw scores were converted to percent correct and then to rationalized arcsine units (RAU). This transformation "stretches" out the upper and lower ends of the scale, thereby allowing for valid comparisons of differences across the entire range of the scale (Studebaker, 1985). Scores on this scale range from −23 RAU (corresponding to 0% correct) to +123 RAU (corresponding to 100% correct).

III. RESULTS

As expected, higher SNRs yielded better target sentence perception in all conditions. Comparison across conditions also shows that sentence perception was better in two-talker noise than in six-talker noise as predicted by previous research (Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005). With respect to the language of the noise, perception was significantly better in Mandarin than in English noise in two-talker babble at SNRs of 0 and −5 dB where those SNRs comprised the second half of the condition. The results for all experimental blocks and conditions are presented in Fig. 1.

Three-way repeated measures analysis of variance (ANOVAs) were performed separately for each condition, with language background (monolingual versus bilingual) as a between-subjects factor, and noise level (easy versus hard SNR) and noise language (English versus Mandarin) as within-subjects factors. There was no main effect of language background, nor any two- or three-way interactions with language background in any of the conditions; therefore language background was removed from all future analyses. This finding established that the mono- and bilingual participants performed equivalently in this study.

Two-way ANOVAs with noise level (easy versus hard) and noise language (English versus Mandarin) as within-subjects factors showed a significant main effect of level in all conditions (condition 1 $[F(1,15)=156.81,\ p<0.0001]$; condition 2 $[F(1,16)=976.09,\ p<0.0001]$; condition 3 $[F(1,16)=80.27,\ p<0.0001]$; condition 4 $[F(1,15)=103.02,\ p<0.0001]$). Condition 4 also showed a significant main effect of noise language $[F(1,15)=7.74,\ p=0.0140]$. Finally, both conditions 3 and 4 (the conditions that used two-talker babble) showed two-way interactions between noise level and noise language (condition 3 $[F(1,15)=6.151,\ p=0.0246]$; condition 4 $[F(1,15)=24.532,\ p=0.0002]$).

Post hoc pairwise comparisons (paired t tests) of the two-talker babble conditions showed a significant difference between English and Mandarin noise at the "hard" levels for

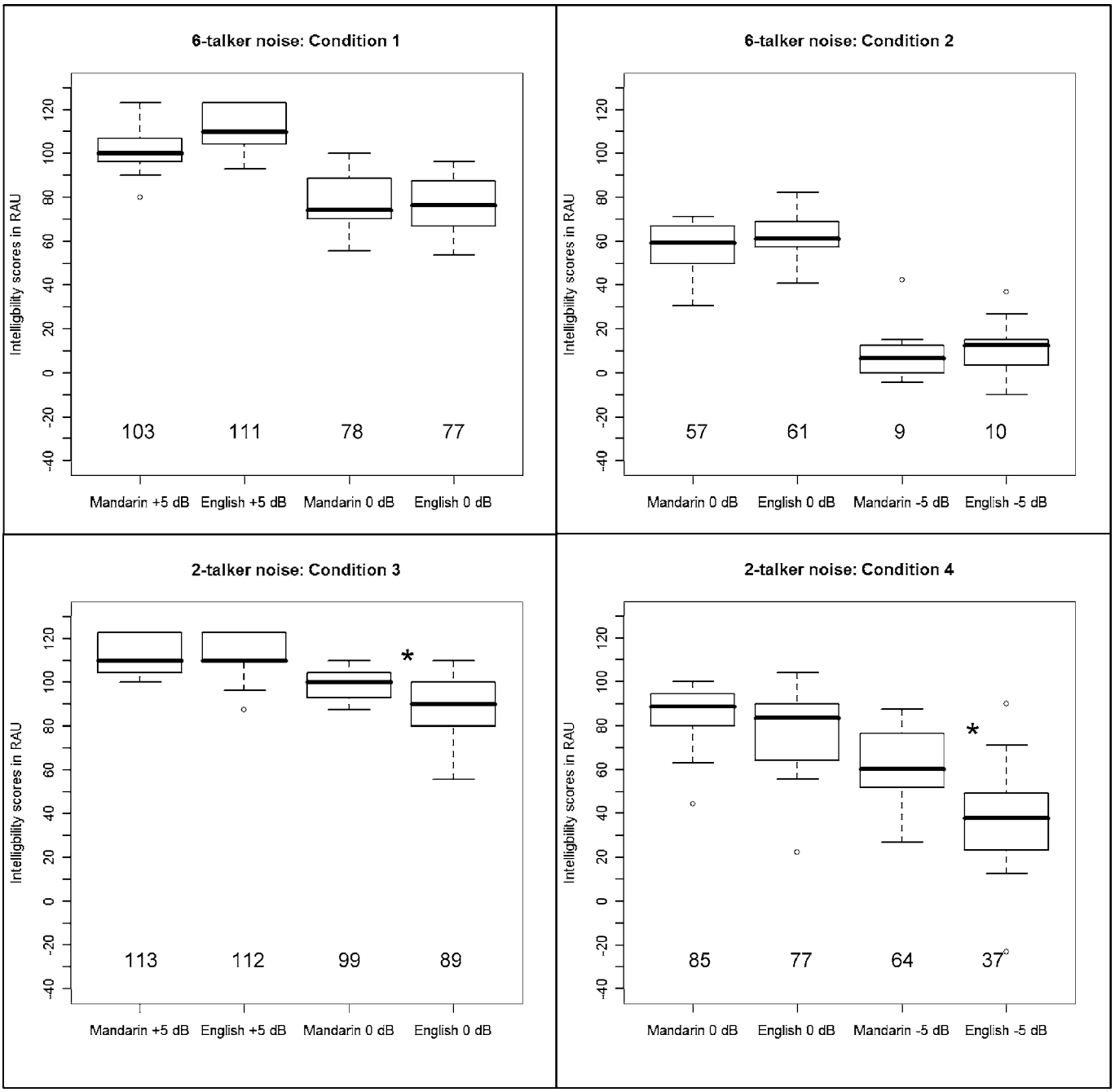

FIG. 1. Boxplots showing the interquartile ranges of intelligibility scores (in RAU) for conditions 1–4. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. Stars indicate significant differences between the two noise languages within a given noise condition. The mean is given at the bottom of each plot.

both condition 3 [$t(16)=2.438$, $p=0.0268$] and condition 4 [$t(15)=4.053$, $p=0.0010$], but no significant difference between the languages at the "easy" level in either condition.

Finally, three-way ANOVAs with noise level (easy vs hard) and noise language (English vs Mandarin) as within-subjects factors and number of noise talkers (two vs six) as a between-subjects factor were performed across conditions 1 and 3 (+5/0 SNRs) and conditions 2 and 4 (0/−5 SNRs). This analysis was included in order to compare the effects of two- vs six-talker noise. Both ANOVAs showed that intelligibility was significantly better in two-talker versus six-talker noise and at higher versus lower SNRs—findings which replicate patterns observed in previous studies (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005).

For conditions 1 and 3, there was a main effect of number of talkers [$F(1,31)=24.65$, $p<0.0001$] and noise level [$F(1,31)=236.43$, $p<0.0001$]. As expected from the analyses on individual conditions, which showed the language effect only for "hard" noise levels and only in two-talker noise, all two-way interactions were significant: noise level and number of talkers [$F(1,31)=12.64$, $p=0.0012$], noise language and number of talkers [$F(1,31)=5.04$, $p=0.0320$] and noise level and noise language [$F(1,31)=6.05$, $p=0.0197$].

For conditions 2 and 4, main effects also emerged for number of talkers [$F(1,31)=61.43$, $p<0.0001$] and for noise level [$F(1,31)=579.40$, $p<0.0001$]. In addition, the effect of language approached significance [$F(1,31)=4.075$, $p=0.0522$]. All two-way interactions were also significant: noise level and number of talkers [$F(1,31)=34.37$, $p<0.0001$], noise language and number of talkers [$F(1,31)=10.23$, $p=0.0032$], and noise level and noise language [$F(1,31)=14.39$, $p=0.0006$]. Finally, there was a three-way interaction between noise level, noise language, and number of talkers [$F(1,31)=9.32$, $p=0.0046$]. Again, these interactions are predicted by the results of the analyses of individual conditions, which showed the effect of language to be significant only in two-talker noise at difficult SNRs. In summary, statistical analysis across two- and six-talker noise conditions indicates that sentence intelligibility is better with fewer talkers in the noise and with lower noise levels.

It should be noted that experience with the task and/or the target talker had an effect on participants' intelligibility scores. This is best illustrated by observing the results from the 0 SNR blocks. Where the first pair of blocks presented to participants was at an SNR of 0 (the "easy" SNR for the condition), they performed worse than when the same SNR was presented in the second half of the experiment (as the "hard" SNR). This difference due to ordering is observed for six-talker noise (compare conditions 1 and 2), as well as for two-talker noise (compare conditions 3 and 4). It is also true for both English and Mandarin noise, and in fact, the language effect only emerged as significant where 0 SNR was in the second half of the two-talker experiment. It is assumed that adjustment to the task and/or increased familiarity with the target voice accounts for such differences. Furthermore, these practice effects outweigh any effects of fatigue, which would cause decline in participant performance over the course of the experimental blocks.

IV. DISCUSSION

This study has shown that the language of interfering noise *can* affect the intelligibility of the target speech (for a similar result, see Rhebergen *et al.*, 2005; Garcia Lecumberri and Cooke, 2006). Specifically, where the noise contains two talkers and is presented at levels equal to or greater than the target speech, English noise is more detrimental than Mandarin noise to native English speakers who are listening to a native English target. This effect of noise language provides evidence that, under certain conditions, linguistic interference plays a role in the perception of speech in noise. The results of this study also replicate previous findings that the perceptibility of speech in multi-talker babble decreases as the number of talkers in the noise increases, as well as by an increase in the level of the noise with respect to the target (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005; Simpson and Cooke, 2005). The overall results are summarized in Table III in the terms laid out in the introduction in Table I.

Before discussing the potential sources of the language effect in the two-talker condition, the lack of this effect in six-talker babble must be addressed. First, it should be noted that in post-experiment interviews, most participants in the six-talker babble conditions did not report noticing a shift in background noise language, and those that were aware of the presence of two noise languages did not report greater distraction from either one. In addition, whereas some subjects transcribed words from the babble in two-talker English conditions, this did not occur in six-talker conditions. The higher intelligibility scores in two- versus six-talker babble overall shows that the greater spectral and temporal density of six-talker noise, which yields greater energetic masking of the signal, made target intelligibility worse. At the same time, these characteristics of six-talker noise eliminated any informational masking differences between the two noise languages. Whatever "benefit" linguistic differentiation between the target and the noise may have provided in two-talker noise conditions was eliminated in six-talker noise. In sum, differential linguistic effects emerged only where the linguistic content of the noise was relatively available to the listener.

TABLE III. Aspects of native versus foreign-language two- and six-talker babble and their predicted effects on target speech intelligibility, presented with intelligibility scores averaged across participants and (matched) conditions.

	Same language two-talkers	Different language two-talkers	Same language six-talkers	Different language six-talkers
Temporal gaps	Yes	Yes	No	No
Linguistic differentiation	No	Yes	No (?)	Yes
Average words correct (RAU)	79	90	65	62

There are several possible sources of the greater masking by English noise than Mandarin noise for English target speech in two-talker babble: first, it is possible that differences in the long-term average spectra of the two languages contributed to the different effects, such that language-specific spectral similarities between the English noise and the English target increased the amount of energetic masking due to greater spectral overlap.

Running *t*-test analysis of the long-term average spectra of the English and Mandarin two-talker noise (averaged across the four tracks in each language) reveals statistically significant differences at several, but not all, frequencies. While these differences may contribute to the intelligibility asymmetry, the differences between the spectra are small and not consistent over the entire spectrum. The overall similarity in the long-term average spectra of the noise in the two languages suggests that spectral differences are not the sole source of the language effect. Furthermore, findings by Byrne *et al.* (1994) showed that, when averaged across talkers, languages do not differ significantly with respect to long-term average spectrum. A more likely explanation for the language effect observed here is that it is indeed an effect of different amounts of informational masking, i.e., a linguistic effect.

The precise aspect(s) of linguistic content that contribute to the language effect remain to be determined. The effect

may be primarily a whole-word lexical effect such that hearing and activating English words in the babble is what makes English noise more difficult to tune out than Mandarin noise. Participants frequently transcribed entire words from the English noise in their responses, indicating that interference occurred at this level. However, there may also be different amounts of interference from noise in different languages at sublexical and/or prosodic processing levels. Differences in the phoneme inventories and syllable structures of the languages, for example, may contribute to differential interference effects, with English phonemes and phonotactic patterns creating greater interference for the English listeners. Differences in rhythmic properties which correlate with syllable structure (Ramus *et al.*, 1999; Grabe and Low, 2002) and prosodic patterns may also contribute to the effect. It is likely that a combination of some or all of these whole-word, sublexical, and prosodic factors contribute to the different amounts of linguistic masking at play when the language of the noise matches or mismatches the language of the target speech. Future research must address their precise contributions to the noise language effect.

One method for conducting a fine-grained analysis of the locus of the linguistic masking reported in this study is to compare English noise to English nonword noise (composed of words that are phonologically legal in English but are not real words). This comparison should allow us to determine whether whole words from the background noise intrude on the target speech or whether sublexical properties can cause as much intrusion as lexical items. Preliminary findings from a study using nonword noise constructed by altering onsets, codas, or vowels in the content words of the original noise sentences showed no significant difference in intelligibility scores. This finding provides evidence against a strict lexical explanation for the language effect. However, it is likely that the high degree of similarity between the nonwords and real English words still causes listeners to activate items in the lexicon. The observation that participants in this preliminary experiment frequently transcribed real words that sounded like the nonwords present in the babble supports this interpretation. In future research, additional manipulations of noise content will be required to provide further insight into the question of the source of the language effect. These may include other types of nonwords, various accents of the target language, babble constructed from nonsentential materials (e.g., syllable strings, word lists), and noise from other languages.

V. SUMMARY/CONCLUSION

This study has shown that higher noise levels and greater numbers of talkers in multi-talker babble decrease target speech intelligibility, no matter what language is being spoken in the noise. However, in certain conditions (few talkers, difficult SNRs), linguistic effects appear to come into play, as shown by the differences observed between the effects of English and Mandarin two-talker babble: native English listeners performed better on a sentence intelligibility task in the presence of Mandarin two-talker babble than in English two-talker babble. We conclude that greater similarity between target and masker in the linguistic domain creates greater interference in target intelligibility, and must be taken into consideration in a principled account of speech perception in noise.

ACKNOWLEDGMENTS

This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD. The authors gratefully acknowledge the help of Ethan Cox.

[1]Simpson and Cooke (2005) showed that this relationship is nonmonotonic. This study varied the number of talkers (N) in both natural talker babble and babble-modulated noise (speech-shaped noise modulated by the envelope of N-talker babble) for a consonant identification task. In natural babble, intelligibility scores decreased with increasing numbers of talkers from N=1 to N=6, but were constant for N=6 to N=128. In babble-modulated noise, by contrast, intelligibility decreased gradually with increasing N values.

[2]The numbers of talkers and SNRs were chosen based on the authors' intuitions, the parameters used in previous speech-in-noise studies, and pilot data. It is recognized, however, that these are, to some extent, arbitrary decisions necessitated by limitations of the experimental design. It remains for future research to investigate a fuller range of talker numbers and SNRs.

[3]It is typically assumed that listeners are able to take advantage of temporal gaps in noise energy to hear relatively unobstructed "glimpses" of the target signal when there are fewer talkers in the noise (Bronkhorst, 2000; Brungart *et al.*, 2001; Assman and Summerfield, 2004; Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2005; Simpson and Cooke, 2005).

[4]The "no" under six-talker, same language noise is marked with a "?" in recognition of the fact that the presence of a greater number of talkers interferes with the availability of linguistic information in the babble due to temporal and spectral overlap of the various speakers in the babble. This babble-internal masking may decrease the linguistic interference that the matched-language noise would otherwise be predicted to impose on the target. However, it is possible that some isolated words and/or other linguistic characteristics will still be identifiable from the six-talker babble, and therefore we mark this cell with "no?" in the table.

Assman, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," *Speech Processing and the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, Berlin), Vol. 18.

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **114**, 1600–l610.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. 86, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3138.

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**, 2527–2538.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., and Wilbraham, K. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Cutler, A., Webber, 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.

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**, 2445–2454.

Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the Rhythm Class Hypothesis," *Papers in Laboratory Phonology 7* (Moutons-Gravenhage).

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. **134**, 477–500.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language

acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.
Nábělek, A. K., and Donohue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.
Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **7**, 265–292.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **114**, 2181–2192.
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 (L)," J. Acoust. Soc. Am. **118**, 1274–l277.
Rogers, C. L., Dalby, J., and Nishi, K. (**2004**). "Effects of noise and proficiency on intelligibility of Chinese-accented English," Lang Speech **47**, 139–154.
Simpson, S. A., and Cooke, M. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.
Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.
Sperry, J. L., Wiley, T. L., and Chial, M. R. (**1997**). "Word recognition performance in various background competitors," J. Am. Acad. Audiol **8**, 71–80.
Studebaker, G. A. (**1985**). "A "Rationalized" Arcsine Transform," J. Speech Hear. Res. **28**, 455–462.
Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.
Van Wijngaarden, S., Steeneken, H., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.