JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0001002

Mandarin tone perception in multiple-talker babbles and speech-shaped noise

Xianhui Wang and Li Xu[a)]

Communication Sciences and Disorders, Ohio University, Athens, Ohio 45701, USA
xw659217@ohio.edu, xul@ohio.edu

Abstract: Lexical tone recognition in multiple-talker babbles ($N = 1$, 2, 4, 8, 10, or 12) and in speech-shaped noise at different signal-to-noise ratios (SNRs = −18 to −6 dB) were tested in 30 normal-hearing native Mandarin-speaking listeners. Results showed that tone perception was robust to noise. The performance curve as a function of N was non-monotonic. The breakpoint at which the performance plateaued was $N = 8$ for all SNRs tested with a slight improvement at $N > 8$ at −6 and −9 dB SNR.

[Editor: Martin Cooke] Pages: EL307–EL313

Received: 15 January 2020 **Accepted:** 12 March 2020 **Published Online:** 3 April 2020

1. Introduction

Multiple-talker babbles (MTBs) and speech-shaped noise (SSN) have been widely used as maskers in studies on speech perception in noise. Two types of masking were suggested to explain the differences in the masking effects of different types of maskers: (1) energetic masking (EM) that derives from the inaudibility of the target due to the overlap of energy between the target and masker across time and frequency regions and (2) informational masking (IM) that arises from the similarity between the target and noise when both are audible and includes many aspects that are beyond peripheral processing (e.g., Brungart *et al.*, 2001).

The total masking effect in MTB is a mixture of EM and IM and is highly dependent on the number of talkers (N) in the babble. As N increases, EM increases monotonically due to the reduced chance for dip-listening [i.e., listeners making use of the temporal-spectral regions with favorable signal-to-noise ratios (SNRs) to extract information of the target of interest]. Recent evidence suggested that EM essentially arises from the interaction of modulation between target and masker rather than their energy in spectro-temporal domain (Stone *et al.*, 2012). Consequentially, dip-listening results from the release from such modulation masking (Stone and Canavan, 2016). EM does not increase anymore after reaching a large-enough N because the energy (modulation) in the temporal-spectral domain saturates. This case can be approximated by SSN that is generated by modulating a random noise with the long-term average spectrum of the speech. IM peaks at a much smaller N (i.e., 2 or 3 talkers; Brungart *et al.*, 2001; Freyman *et al.*, 2004) and has a much more profound impact on speech intelligibility than EM (Brungart *et al.*, 2006). Many important aspects of IM have been demonstrated in previous studies such as the speaker-masker gender (i.e., voice information coded in the pitch) (Brungart *et al.*, 2001), the relative spatial location between the target and masker (Freyman *et al.*, 1999, 2004; Rakerd *et al.*, 2006), and the acoustic-phonetic masking and the lexical interference (Hoen *et al.*, 2007). IM barely contributes to the overall masking when N is greater than 128 (Simpson and Cooke, 2005).

Using various speech materials as targets, many studies have investigated perceptual performance in babble maskers. Recognition performance was almost always a non-monotonic function of N, that is, a rapid decrease to its minimum at a relatively small N followed by a plateau with gradual improvement as N becomes greater. Rosen *et al.* (2013) addressed this common finding among separate studies and named the N separating the performance curve into different trends (i.e., slopes) as a "breakpoint." At the breakpoint, the performance was usually the poorest. In Miller (1947), the breakpoint in masking isolated English words was at $N = 6$. Carhart *et al.* (1975) observed that the perception of spondees was worst at $N = 3$. The breakpoint in Rosen *et al.* (2013) was at $N = 2$ where natural English sentences were used as targets. The perception of vowel-consonant-vowel (VCV) syllables yielded a breakpoint at $N = 8$ (Simpson and Cooke, 2005). Although the breakpoint varied among target materials, an improvement in performance at N greater than the breakpoint was almost always present at −6 dB SNR. Rosen *et al.* (2013) attributed the improved performance to the release of the overwhelming effects of IM.

[a)]Author to whom correspondence should be addressed.

Nevertheless, most of the investigation of speech perception under MTB was conducted in non-tonal languages such as English. It remains unknown if this would also be the case for lexical tones. Mandarin Chinese has four distinct lexical tones that are typically termed as Tone 1 (high and flat), Tone 2 (rising), Tone 3 (falling and then rising), and Tone 4 (falling). Lexical tones function just as vowels and consonants do in non-tonal languages, that is, to discriminate the meaning of the monosyllable. The perception of lexical tones has been found to be relatively robust to degraded listening environments (Xu *et al.*, 2002; Kong and Zeng, 2006; Krenmayr *et al.*, 2010; Lee *et al.*, 2013; Qi *et al.*, 2017). According to Qi *et al.* (2017), for example, the tone perception was approximately 60% correct at −12 dB SNR in SSN and 55% correct at −18 dB SNR in two-talker babbles. These results are expected given that voice contrasts are relatively robust in noise for non-tonal language (Miller and Nicely, 1955).

The purposes of the present study were to compare the psychometric functions of tone recognition in SSN and MTB and to determine how lexical tone perception in MTBs is affected by *N*. The majority of studies on English speech perception in noise used SNRs that were greater than −12 dB in order to avoid floor performance (Miller, 1947; Brungart *et al.*, 2001; Freyman *et al.*, 2004; Simpson and Cooke, 2005; Rosen *et al.*, 2013). Based on Qi *et al.* (2017) and our pilot experiments, we chose −18 to −6 dB SNRs in the present study to avoid floor or ceiling effects in tone perception performance. To determine the performance of lexical tone perception in babbles as a function of *N* and to compare that to the existing results of studies on the English language, we examined the intelligibility of lexical tones in babbles consisting of 1, 2, 4, 8, 10, and 12 talkers and in SSN.

2. Methods

2.1 Subjects

Thirty native-Mandarin-speaking adults (15 males and 15 females) were recruited to participate in the study. All listeners were university undergraduate or graduate students in Beijing. Most of them were Beijingese who speak standard Mandarin Chinese and Beijing dialect. A few of them were from other provinces but they had stayed in Beijing for more than 3 years and were fluent in standard Mandarin Chinese. The age of the participants ranged from 18 to 45 years [25.3 ± 3.1, mean and standard deviation (s.d.)]. All participants were screened for normal hearing (≤20 dB hearing level) at octave frequencies between 250 and 8000 Hz. No participant had any history of speech or hearing disorders. The use of human subjects was reviewed and approved by the Institutional Review Board of Ohio University.

2.2 Materials

The stimuli for the tone recognition test consisted of ten monosyllables: /fu/ "fu," /tɕi/ "ji," /ma/ "ma," /tɕʰi/ "qi," /uan/ "wan," /ɕi/ "xi," /ɕiɛn/ "xian," /iɛn/ "yan," /iaŋ/ "yang," and /i/ "yi," with each of them being in four tones. The tokens were recorded from one male and one female native Mandarin speaker. The mean *F*0 for the male and female speakers were 137 and 257 Hz, respectively. In total, there were 80 tone tokens (10 syllables × 4 tones × 2 speakers). The durations of the four tones of each syllable were equalized to the mean duration of the four tones of each syllable using the method of the pitch-synchronous overlap-add method (Boersma and Weenink, 2016). All tone tokens were then adjusted to the same root-mean-squared (RMS) amplitude and an interval of 300-ms silence was added both before and after the duration-equalized tone token.

The SSN was generated by filtering a white noise to the long-term average speech spectrum of the 80 tone tokens used in the present study. To generate MTB, narrative speech in Mandarin Chinese was recorded from 12 native speakers of Mandarin Chinese (6 females and 6 males). The speakers were all broadcasting major undergraduate seniors aged between 22 and 24 years old. The mean *F*0s of the 6 male speakers were 93, 99, 103, 107, 124, and 148 Hz and those of the 6 female speakers were 138, 159, 196, 210, 212, and 253 Hz. The number of talkers (*N*) for MTB included 1, 2, 4, 8, 10, and 12. A particular number of narrative speech segments (corresponding to *N*) with an appropriate length was randomly selected from the 12 possible speech recordings, some from the females and the rest from the males. For the even number of talkers (i.e., 2, 4, 8, 10, and 12), equal numbers of female and male talkers were selected randomly. For the odd number of talkers (i.e., 1), a female or a male talker was selected randomly. The RMS amplitude of those speech segments from different talkers was equalized before use.

On each presentation, the SSN or MTB with appropriate length was mixed together with the tone token at a specific SNR (i.e., −18, −12, −9, or −6 dB). The RMS amplitude of tone tokens was fixed and the RMS of the SSN or MTB was manipulated to achieve the desired SNRs. The masker was present 300 ms before the tone tokens and 300 ms after the end of the

https://doi.org/10.1121/10.0001002

tone tokens. A 50-ms cosine ramp was applied to the onset and offset of the target-masker mixture.

2.3 Test procedure

The tone-masker mixtures were presented binaurally to participants through Sennheiser HD280 Professional headphones (Sennheiser, Wedemark, Germany) in a double-walled sound booth with noise levels $\leq$30 dB (A). The intensity of presentation was adjusted to the participant's most comfortable level. A custom MATLAB program was used to conduct the tone recognition test. Before the formal testing, a practice session was conducted to allow familiarization of the testing procedure and stimuli.

The test was conducted using a four-alternative forced-choice procedure, in which four Chinese characters with the same syllable and the PINYIN with tonal marking were displayed on the screen and the participants were asked to select one out of the four tones based on what they heard. There were a total of 28 test conditions [i.e., (1 SSN + 6 MTB) $\times$ 4 SNRs]. For each participant, the order of the 28 test conditions was randomized. In each condition, 80 tokens (10 syllables $\times$ 4 tones $\times$ 2 speakers) were used and the order of tone token presented was also randomized. Each participant completed a total of 2240 trials (80 tokens $\times$ 28 conditions). The test took approximately 60 to 80 min to complete.

2.4 Data analyses and statistical methods

Data analysis was performed in MATLAB with the Statistics Toolbox. The percent-correct scores of the tone-perception test were binomial data and thus were analysed using a generalized linear mixed model (GLMM) (Warton and Hui, 2011) to examine the effects of (1) the number of talkers (N), (2) SNRs, and (3) the interaction between N and SNR. The number of correct responses out of 80 for each condition for each listener was entered into the GLMM model. The Pearson chi-square dispersion statistic (the ratio of the Pearson chi-square statistic to its degrees of freedom) was calculated to evaluate dispersion in GLMM. Planned comparisons on N were conducted at each SNR separately. Specifically, the performance at each N to that of SSN for a particular SNR were paired and compared. The Bonferroni correction was used to control the familywise type I error rate at 0.05.

3. Results and discussion

Figure 1 shows the tone-recognition performance as a function of SNR. The Pearson chi-square dispersion statistic in GLMM was 1, indicating that no overdispersion was detected. The GLMM analyses revealed significant main effects of SNR and the type of maskers (both $p < 0.0001$). The two-way interaction between SNR and the type of noise was also significant ($p < 0.0001$), indicating that the impact of the type of masker on tone recognition changed as the SNR varied. Specifically, while tone perception performance was fairly good at -6 dB SNR for all maskers,

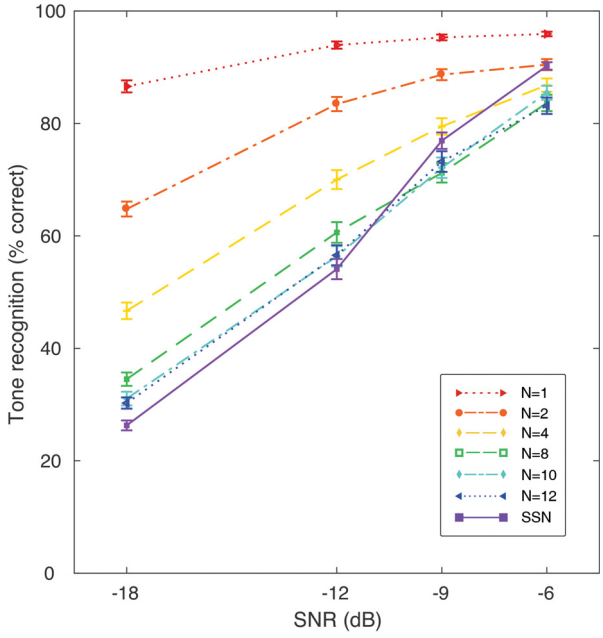

Fig. 1. (Color online) Group mean tone-recognition scores as a function of SNR. Each line represents the recognition performance in SSN or MTB of a specific N. Error bars indicate $\pm$1 standard error (S.E.).

decreases in SNR led to much greater decrements in performance for higher numbers of talkers in the babble (or SSN), in comparison to babbles with fewer talkers. This is also reflected by the steeper slope of tone recognition performance as a function of SNR at a greater N in the babble.

These results demonstrated that lexical tone perception was robust to both MTB and SSN. A previous study showed that speech reception thresholds (SRTs, the minimum SNR required to reach 62.5% correct) of Mandarin tones in SSN were, on average, between −13.6 and −12.9 dB SNR (Krenmayr *et al.*, 2010). From Fig. 1, it can be estimated that the SRTs for SSN from the present study was consistent with those reported in Krenmayr *et al.* (2010). The percent-correct scores in SSN were also comparable to those reported in Qi *et al.* (2017). In comparison to one MTB condition (i.e., $N=2$) that was tested in Qi *et al.* (2017), the present study showed a slightly higher score of about 5–10 percentage points at comparable SNRs. The reason might be due to the fact that we used natural tone tokens in the present study whereas in the Qi *et al.* (2017) study, the tone tokens went through the chimera processing and were resynthesized. This process might have caused decreased recognition performance in noise conditions.

The sex of the speakers of the tone tokens showed a significant effect on tone-recognition performance. Figure 2 displays the mean tone-recognition scores separated by the sex of speakers of the tone tokens in various types of maskers. Results from all SNRs (i.e., −18, −12, −9, and −6 dB) were pooled together. The paired t-test indicated that on average, the tone recognition score was higher when the speaker was a female than it was a male for all MTB conditions (all $p<0.0001$) and for the SSN condition ($p<0.05$). Averaged across all masking conditions and SNRs, tone recognition using female tone tokens was 16.6 percentage points higher than that using male tone tokens.

Although it is difficult to draw solid conclusions about the difference in performance relating to the sex of speakers given the limited number of speakers used in the present study (one male and one female), many studies in non-tonal languages have reported similar results in speech perception in noise. Krenmayr *et al.* (2010) showed that tone recognition in SSN with the female speaker yielded a lower SRT than with the male speaker. The SRTs were −13.6 and −12.9 dB SNR for the female and male speakers, respectively. Female voice is different from male voice in many acoustic parameters such as $F0$ (Fry, 1979), formant frequencies (Peterson and Barney, 1952), and the degree of acoustical contrasts (Koopmans-van Beinum, 1980). English-speaking females were reported to use rising intonation more than males (Jiang, 2011). Although the relationship between these parameters and speech intelligibility is still under debate, studies on non-tonal languages showed that the female voice produced higher speech intelligibility scores than the male voice (Kwon, 2010; Bradlow *et al.*, 1997; Byrd, 1994). However, the influence of these parameters on lexical tone perception in noise was not clear. Presumably, a female voice has a higher $F0$ than a male voice and thus a greater separation of higher harmonics. Such spaced-out harmonics might be easier to be resolved in the auditory system in noise conditions, which in turn might facilitate lexical tone recognition. The effects of speaker sex or voice $F0$ on tone recognition in noise will be comprehensively investigated in our future research.

Figure 3 displays tone-recognition scores as a function of N of the MTB. For comparison, we also plotted in Fig. 3 the recognition performance of English VCV syllables (Simpson and Cooke, 2005), words (Miller, 1947), and sentences (Rosen *et al.*, 2013). Tone-recognition scores of the present study at −6, −9, and −12 dB SNRs were better than the perception scores measured using English VCV syllables, words, or sentences at −6 dB SNR. At −18 dB SNR,

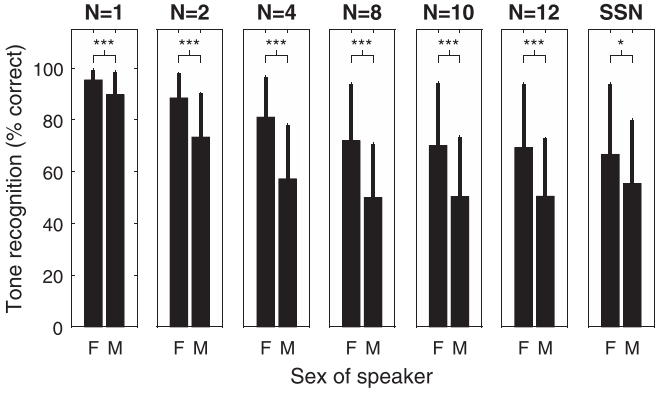

Fig. 2. Group mean tone-recognition scores separated by the sex of speakers at different N of the MTB and in SSN. Data were pooled from all SNRs (i.e., −18, −12, −9, and −6 dB). Error bars indicate 1 s.d., *** represents a significance level of $p<0.0001$, and * $p<0.05$.

JASA EXPRESS LETTERS
https://doi.org/10.1121/10.0001002

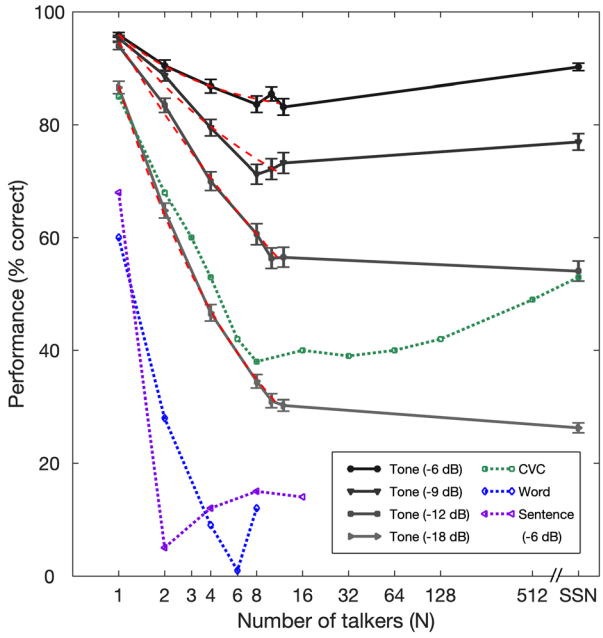

Fig. 3. (Color online) Group mean speech-recognition scores as a function of N of MTBs. Recognition performance in SSN is arbitrarily plotted beyond $N = 512$. Solid lines with filled symbols are tone-recognition scores at different SNRs in the present study. Error bars indicate ±1 S.E. Dashed lines without symbols are fitted curves from the exponential regression. Dotted lines with unfilled symbols are speech-recognition scores with different speech materials from previous studies on the English language: VCV syllables (Simpson and Cooke, 2005), words (Miller, 1947), and sentences (Rosen *et al.*, 2013).

tone recognition scores were slightly poorer than that of English VCV syllables at −6 dB SNR but were still better than those of other English speech materials. Note that the chance performance in tone recognition task is 25% (1/4), which is much higher compared to the chance performance in consonant perception (1/16, about 6.3%) and word and sentence perception that are open-set tests.

Next, the tone recognition scores as a function of N of MTBs were fit with an exponential regression function, $y = ae^{b(n+c)} + d$, in which y is the percentage correct scores and n is the number of talkers that was used to fit the group mean score curves at each SNR, separately. The values of the parameters of the exponential function were determined based on the ordinary least square method (Xu and Zheng, 2007). The breakpoint for each of the fitting curves was defined as the number of talkers at which 90% of the performance plateau is reached. Note that our method of calculating breakpoint was different from that used in Rosen *et al.* (2013). In the latter, the breakpoint was defined as the number of talkers at which slopes of two fitted straight lines of the performance data differed the most. The breakpoint was found at approximately eight talkers for all SNR conditions in the present study. The improvement of tone-recognition performance after the breakpoint occurred only in the conditions where SNR was more favorable (i.e., −6 and −9 dB) but not in the least favorable conditions (i.e., −12 and −18 dB).

An overall good performance was observed at $N = 1$ and 2 (see also Fig. 1). The effects of IM and EM were minimum for these values of N. Brungart (2001) and Brungart *et al.* (2001) suggested that listeners were able to tack the voice of speaker when $N = 1$. It is likely that our listeners also used such pitch information of target speakers (i.e., one male and one female) to track target tones. The rapid decrease in performance when N was <8 was monotonic. Apart from the monotonic increase in EM, this may also be addressed by one of the aspects in IM: the acoustic-phonetic masking, which refers to the masking effect due to the presence of acoustic-phonetic information (such as acoustic correlates of phonemes) in the background babble. Note that acoustic-phonetic masking is independent of the meaningfulness (or lexicality) of maskers. A timely-reversed speech babble (i.e., not meaningful) has the same amount of acoustic-phonetic masking as a natural speech babble (meaningful) when they have the same N (Hoen *et al.*, 2007). As N increases from 1 to 8, the acoustic-phonetic information in babbles increases monotonically (Hoen *et al.*, 2007). Such acoustic-phonetic information in babbles of the present study was likely to be the acoustic correlates of lexical tones: the $F0$ contour and amplitude contours (Whalen and Xu, 1992). Presumably, another aspect of IM contributing to the rapid decrease in performance might be the distraction due to the background lexicality (i.e., being meaningful). The lexicality in the background causes failure in object selection that is directed by a top-down process

(Shinn-Cunningham *et al.*, 2007). In the present study, the lexical information in babbles may have preoccupied listeners' attention over a single tone target.

The breakpoint occurred at approximately eight talkers in all SNR conditions (see Fig. 3). This was comparable with that observed in English VCV perception (Simpson and Cooke, 2005), but was greater than those in English word recognition (Miller, 1947) and sentence recognition (Rosen *et al.*, 2013). Perhaps the amount of IM is different in situations whether the targets activate the semantic processing (as in word and sentence recognition) or not (as in VCV and tone recognition).

The improvement of tone recognition performance on the plateau only occurred at better SNR conditions (i.e., −6 and −9 dB), but not at poorer conditions (i.e., −12 and −18 dB). This may imply a shift of the role that IM plays as SNR changes, that is, IM is stronger when SNR is more favorable. Similarly, Brungart (2001) observed the worst performance resulting from gender similarity at a positive SNR but not a negative SNR. Considering conditions where IM is stronger, the release from IM due to the increment of N should have greater magnitude. Therefore, our result is consistent with the notion of Rosen *et al.* (2013) in that the improvement in performance occurred when the release from IM overwhelms EM that monotonically increases with N under more favorable SNR conditions.

In summary, the present study demonstrated the robustness of lexical tones under SSN and MTB consisting of $N = 1$, 2, 4, 8, 10, and 12. Under all masking conditions, tone recognition using female tone tokens was higher than that using male tone tokens. Across all masking conditions and SNRs tested (i.e., −18, −12, −9, and −6 dB), tone recognition was approximately 16.6 percentage points higher with the female voice than with the male voice used in this study. Tone-recognition performance as a function of N was non-monotonic in various SNRs tested. The breakpoint observed in our study was $N = 8$, similar to that in English VCV syllable recognition reported by Simpson and Cooke (2005) but greater than those reported in English word and sentence recognition (Miller, 1947; Rosen *et al.*, 2013). The improvement of performance after breakpoint was only observed in better SNR conditions but not in the poorer SNR conditions.

Acknowledgments

This study was partially supported by a grant from the NIH/NIDCD (Grant No. R15-DC014587). Lexi Neltner provided editorial assistance in the preparation of the manuscript.

References and links

Boersma, P., and Weenink, D. (**2016**). "PRAAT: Doing phonetics by computer (version 6.0.21) [computer program]," http://www.praat.org.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1997**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

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

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

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.

Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **15**, 39–54.

Carhart, R., Johnson, C., and Goodman, J. (**1975**). "Perceptual masking of spondees by combinations of talkers," J. Acoust. Soc. Am. **58**(S1), S35.

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**, 2246–2256.

Freyman, R. L., Helfer, K. S., Mccall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Fry, D. B. (**1979**). *The Physics of Speech* (Cambridge University Press, Cambridge, England).

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

Jiang, H. (**2011**). "Gender difference in English intonation," in *International Congress of Phonetic Sciences*, Melbourne, Australia (August 4–10), pp. 974–977.

Kong, Y.-Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.

Koopmans-van Beinum, F. J. (**1980**). *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions* (Academische Pens, Amsterdam).

Krenmayr, A., Qi, B., Liu, B., Liu, H., Chen, X., Han, D., Schatzer, R., and Zierhofer, C. M. (**2010**). "Development of a Mandarin tone identification test: Sensitivity index d' as a performance measure for individual tones," Int. J. Audiol. **50**, 155–163.

Kwon, H. B. (**2010**). "Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses," J. Adv. Prosthodont. **2**, 71–76.

https://doi.org/10.1121/10.0001002

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2013**). "Effects of speaker variability and noise on Mandarin tone identification by native and non-native listeners," Speech Lang. Hear **16**, 46–54.
Miller, G., and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**(2), 338–352.
Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Qi, B., Mao, Y., Liu, J., Liu, B., and Xu, L. (**2017**). "Relative contributions of acoustic temporal fine structure and envelope cues for lexical tone perception in noise," J. Acoust. Soc. Am. **141**, 3022–3029.
Rakerd, B. L., Aaronson, N. M., and Hartmann, W. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (**2013**). "Listening to speech in a background of other talkers: Effects of talker number and noise vocoding," J. Acoust. Soc. Am. **133**, 2431–2443.
Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (**2007**). "A sound element gets lost in perceptual competition," Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227.
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.
Stone, M. A., and Canavan, S. (**2016**). "The near non-existence of 'pure' energetic masking release for speech: Extension to spectro-temporal modulation and glimpsing," J. Acoust. Soc. Am. **140**(2), 832–842.
Stone, M. A., Fullgrabe, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am. **132**(1), 317–326.
Warton, D. I., and Hui, F. K. (**2011**). "The arcsine is asinine: The analysis of proportions in ecology," Ecology **92**, 3–10.
Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.