ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Intelligibility of Noise-Adapted and Clear Speech in Energetic and Informational Maskers for Native and Nonnative Listeners

Kirsten Meemann[a] and Rajka Smiljanić[a]

[a] Department of Linguistics, The University of Texas at Austin

ARTICLE INFO

Article History:
Received March 25, 2021
Revision received October 19, 2021
Accepted December 11, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Stephanie Borrie

https://doi.org/10.1044/2021_JSLHR-21-00175

ABSTRACT

Purpose: This study explored clear speech (CS) and noise-adapted speech (NAS) intelligibility benefits for native and nonnative English listeners. It also examined how the two speaking style adaptations interact with maskers that vary from purely energetic to largely informational at different signal-to-noise ratios (SNRs).

Method: Materials consisted of 40 sentences produced by 10 young adult talkers in a conversational and a clear speaking style under two conditions: (a) in quiet and (b) in response to speech-shaped noise (SSN) played over headphones (NAS). Young adult native (Experiment 1) and nonnative (Experiment 2) English listeners heard target sentences presented in two-talker (2T) babble, six-talker (6T) babble, or SSN and at an "easier" and a "harder" SNR.

Results: When talkers produced CS and NAS, word recognition accuracy was significantly improved for both listener groups. The largest intelligibility benefit was obtained for the CS produced in response to noise (CS+NAS). Overall accuracy was highest in 2T babble. Accuracy was higher in SSN than in 6T babble for nonnative listeners at both levels of listening difficulty but only at a more difficult SNR for native listeners. Listeners benefited from CS and NAS most in the presence of SSN and least in 2T babble. When SNRs were the same for the two listener groups, native listeners outperformed nonnative listeners in almost all listening conditions, but nonnative listeners benefited more from CS and NAS in 6T babble than native listeners did.

Conclusions: Combined speaking style enhancements, CS+NAS, provided the largest intelligibility increases for native and nonnative listeners in all listening conditions. The results add to the body of evidence supporting speech-oriented, behavioral therapy techniques for maximizing speech intelligibility in everyday listening situations.

Comprehending speech in complex, realistic, acoustic environments is a challenging task even for a healthy, normal-hearing person. Background noise and competing speech can impede access to the target speech signal (Brungart et al., 2001; Cooke, 2006; Cooke et al., 2008; Festen & Plomp, 1990; Mattys et al., 2009, 2012), and the signal itself can vary in terms of how clearly it is produced. In addition to these signal-dependent challenges, nonnative and bilingual listeners can vary in their experience acquiring the target language, which impacts speech processing and can make speech perception in noise disproportionately difficult (Garcia Lecumberri et al., 2010; Van Wijngaarden et al., 2002). Each of these factors plays a role in daily communication, although the many ways in which they interact still need to be determined. The goal of this study was to extend our understanding of how variation in speech clarity through speaking style modifications aids word recognition in different maskers and noise levels. Additionally, we wanted to compare native and nonnative

Correspondence to Kirsten Meemann: kirsten.meemann@utexas.edu.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

listener performance in these varied listening conditions. This is relevant as daily interactions increasingly occur between native and nonnative interlocutors in restaurants, hospitals, classrooms, and open-plan offices where noise is ubiquitous and can affect the quality of communication, productivity, and health (Lou & Ou, 2020; Peng & Wang, 2016). With better understanding of how speaking style modifications facilitate speech processing in a variety of listening conditions and for different listener groups, specific and ecologically more relevant information can be incorporated into training, educational, and clinical practices, with the goal of improving intelligibility (Godoy et al., 2013; Park et al., 2016; Pichora-Fuller et al., 2010; Tjaden, Kain, & Lam, 2014; Tjaden et al., 2013; Tjaden, Sussman, & Wilding, 2014; Zeng & Liu, 2006).

Talkers typically adjust their speech in response to communication challenges (Cooke et al., 2014; Pichora-Fuller et al., 2010; Smiljanić, 2021; Smiljanić & Bradlow, 2009). Clear speech (CS) and noise-adapted speech (NAS) are two adaptations produced in response to listener limitations (Picheny et al., 1986) and noisy environments (Lombard, 1911), respectively. Here, we use CS to refer to a deliberate, clear speaking style talkers produce when communicating with someone who has hearing loss or low proficiency in the target language. We use NAS to refer to the more automatic speech modifications talkers produce to overcome reduced audibility due to environmental noise. The acoustic characteristics and perceptual benefits of CS and NAS compared with conversational speech (CO) and speech produced in quiet (Q), respectively, have been reported for various talker and listener populations as well as listening conditions (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Cooke & Garcia Lecumberri, 2012; Cooke et al., 2013; Ferguson, 2004, 2012; Hazan & Simpson, 2000; Krause & Braida, 2009; Picheny et al., 1985; Smiljanić, 2013). Only a few studies to date, however, have compared the intelligibility benefit of NAS and CS directly, reporting conflicting results. Goy et al. (2013), for example, found that NAS, not CS, significantly improved intelligibility for older adult listeners in difficult listening conditions (with more background noise) but not for younger adult listeners. In contrast, Smiljanić and Gilbert (2017b) showed that both CS and NAS significantly enhanced intelligibility for young adult listeners in the presence of a speech-shaped noise (SSN) masker, with a larger intelligibility gain for NAS alone than for CS alone. The two speech modifications produced together (clear speech produced under noisy conditions [CS+NAS]) provided the largest intelligibility benefit. Acoustic analyses (Smiljanić & Gilbert, 2017a) revealed that NAS and CS involved similar signal augmentation changes through a decrease in speaking rate and an increase in sound pressure level, vowel space area, and 1- to 3-kHz energy, which characterizes speech production with increased vocal effort (Glave & Rietveld, 1975; Sluijter & van Heuven, 1996). The two speaking styles also showed several differences reflecting the specific type of adverse listening condition speakers were trying to overcome (e.g., increased f_o mean and decreased jitter and shimmer in NAS, increased frequency and duration of pauses in CS; see also Cooke et al., 2014; Cooke & Lu, 2010; Godoy et al., 2013; Goy et al., 2013; Lu & Cooke, 2008, 2009). Here, we build on this work to shed further light on whether the CS and NAS modifications, separately and in conjunction, provide intelligibility benefit in different listening conditions—different maskers and noise levels—and for native and nonnative listeners. One motivation for carrying out this work was the difficulty in comparing findings across research studies as they examined some of these factors separately, differing in speech elicitation methods, speech materials, listening conditions, and listeners.

This article compares the CS and NAS intelligibility benefits in communicative situations typically characterized by acoustic degradation of the target speech signal. The purpose was to determine how perceptual benefit changes through differing amounts of energetic masking (EM) and informational masking (IM) present in background noise (Brungart et al., 2001; Freyman et al., 2004; Pollack, 1975). EM from, for example, steady-state SSN, which is frequently utilized in laboratory and clinical testing, occurs in the auditory periphery as spectrotemporal overlap of the target signal and the noise signal, making portions of the target speech less audible. The amount of EM also depends on the target–masker signal-to-noise ratio (SNR). IM is the additional perceptual difficulty that occurs beyond what can be accounted for by EM (Brungart, 2001; Brungart et al., 2001; Cooke et al., 2008; Durlach, 2006; Freyman et al., 2004). For example, maskers composed of competing speech or speech babble are dominated by IM and can be more detrimental to speech intelligibility than stationary noise, which does not contain informational content (Brungart et al., 2006; Durlach et al., 2003; Freyman et al., 2007; Rosen et al., 2013). On the other hand, speech babble maskers provide "glimpsing" opportunities of the target signal due to fluctuations in masker energy. These spectrotemporal dips provide relief from EM so that the target speech energy exceeds the masker energy, allowing listeners to access the target signal (Cooke, 2006). The number of glimpsing opportunities depends on the number of talkers in the speech masker. With an increase in the number of background talkers (e.g., six talkers vs. two talkers), linguistic interference causes IM while providing fewer glimpsing windows for listeners, thereby increasing EM (e.g., Festen & Plomp, 1990; Freyman et al., 2004; Kilman et al., 2014; Simpson & Cooke, 2005; Van Engen et al., 2014). Finally, speech maskers in listeners' native language have been found to be more detrimental for target speech recognition than those in an unfamiliar language (Garcia Lecumberri & Cooke, 2006; Van Engen, 2010; Van Engen & Bradlow, 2007).

There is some evidence that maskers that vary in the degree of IM and EM impact the intelligibility benefit listeners obtain from CS and NAS. For example, Lu and Cooke (2008) showed that NAS was more intelligible for matrix sentences presented in the same background of EM and IM used to elicit speaking style modifications compared with speech produced in quiet, although the magnitude of the effect was compatible with the EM effect of the competing talker. Hazan et al. (2012) showed that the matched target–masker tokens (speech produced when hearing speech babble and tested in the presence of speech babble) were also identified faster compared with the mismatched target–masker tokens (speech produced in response to vocoded speech and tested in the presence of speech babble), even though all productions were rated as similarly clear. This effect may underlie the larger intelligibility benefit of NAS (produced in response to SSN and tested in the presence of SSN) compared with that of CS (produced in quiet and tested in the presence of SSN) found in the work of Smiljanić and Gilbert (2017b). To our knowledge, the only study that examined the effect of maskers that varied in the degree of EM and IM (two-talker [2T] babble, four-talker babble, eight-talker [8T] babble, and SSN) on CS benefit was that of Van Engen et al. (2014). In contrast to the studies described above, Van Engen et al. found no interactions between masker and speaking style. That is, the CS benefit was constant in the presence of all maskers. In order to tease apart the factors that contribute to the intelligibility benefit the speaking style enhancements provide, it is thus crucial to consider the characteristics of the listening environment in relation to the specific adaptations talkers implement.

Most speech perception studies discussed so far focused on intelligibility variation due to the acoustic signal degradation, such as different maskers, or enhancements, such as CS or NAS. This study also examines intelligibility for nonnative listeners who face the signal-independent challenges of inexperience with the target language and interference from their first language and for whom speech perception in noise remains particularly problematic (Cooke et al., 2008; Cutler et al., 2008; Mayo et al., 1997; Nábělek & Donahue, 1984; Rogers et al., 2006; Takata & Nábělek, 1990). Compared with native listeners, nonnative listeners differ in sensitivity to acoustic cues that are important for mapping the acoustic input to stored representations and extracting the intended meaning from the speech signal. These differences are evident at all levels of linguistic processing: phoneme distinction and interword competition (Cutler et al., 2006; Weber & Cutler, 2004), syntactic processing (Clahsen & Felser, 2006), or prosodic processing (Akker & Cutler, 2003). Garcia-Lecumberri and Cooke (2006), for example, showed that nonnative listeners performed worse than native listeners on consonant identification in SSN, 8T babble, and single-talker competing speech, even though the two listener groups performed similarly in quiet (see also Cutler et al., 2008). The disproportionate effect of noise on nonnative speech perception was also found for word and sentence materials and with increased signal distortion (decreased SNR; Bradlow & Bent, 2002; Cooke et al., 2008; Mayo et al., 1997; Van Engen, 2010). This difficulty is evident even when the proficiency in the target language is high (Brouwer et al., 2012; Rogers et al., 2006; Van Wijngaarden et al., 2002).

Nonnative listeners' ability to take advantage of speaking style modifications for word recognition under different communicative challenges remains largely unexplored. Bradlow and Bent (2002) found a substantially smaller CS gain for low-proficiency nonnative listeners compared with native listeners. Bradlow and Alexander (2007) found that nonnative word recognition in noise was improved only when both acoustic information and semantic information were enhanced, that is, for high-predictability CS sentences relative to low-predictability conversational sentences. In contrast, native adult listeners benefited from semantic context and signal clarity separately and in conjunction. As nonnative listeners gain expertise in target language processing, they increasingly manage to attend to and utilize CS enhancements, although they still need better signal quality (higher SNR) than native listeners for the same intelligibility gain (Smiljanić & Bradlow, 2011). Similarly, even though nonnative listeners benefited from NAS, the benefit was not as large as for native listeners (Cooke & Garcia Lecumberri, 2012; Lu & Cooke, 2008). None of the studies discussed above, however, compared directly how the two speaking style adaptations facilitate native and nonnative listeners' word recognition in different maskers and noise levels.

The goal of this study was to directly compare the intelligibility benefit provided by two speaking style adaptations in response to listener perceptual difficulty (CS) and environmental noise (NAS), separately and in combination. In addition, we were interested in the variation in CS and NAS intelligibility benefits in different listening conditions, that is, in the presence of maskers varying in the amount of EM and IM contributions and at different SNRs. Finally, we wanted to examine whether native and nonnative listeners show similar CS and NAS intelligibility improvements in the presence of different maskers and noise levels. Specifically, we compared word recognition for meaningful sentences produced in conversational and clear speaking styles, first produced in quiet and then in response to SSN. Native (Experiment 1) and nonnative (Experiment 2) listeners heard these sentences in the presence of SSN or overlapping speech from two talkers (2T babble) or six talkers (6T babble) at different SNRs. Table 1 shows the procedure used to elicit the production of CS and NAS (Smiljanić & Gilbert, 2017a) and the listening conditions from Experiments 1 and 2 used in this study. Using the same speech materials will allow us to draw more meaningful conclusions about the CS and NAS intelligibility benefits separately and when produced in combination. Furthermore, the use of meaningful sentences, produced by 10 talkers, will enhance our understanding of the CS and NAS intelligibility variation that was obtained from the studies using, for example, matrix sentences (e.g., Lu & Cooke, 2008) or produced by only one talker (e.g., Van Engen et al., 2014).

On the basis of the research reviewed above, we made several predictions. First, we looked at the effects of speaking style, masker, and SNR as well as their interactions on listeners' word recognition. We examined these effects separately for native (Experiment 1) and nonnative (Experiment 2) listeners. We predicted that (a) CS and NAS will each significantly improve word recognition in noise, and the benefit is expected to be largest when the two modifications are produced in conjunction. (b) Intelligibility will be higher in 2T babble with relatively low EM compared with a purely energetic masker (SSN). Overall accuracy was expected to be lowest in 6T babble due to the combined IM and EM. (c) Intelligibility will decrease as the signal is more degraded (at harder SNRs). (d) We predicted similar but less pronounced main effect results for nonnative as for native listeners. Next, we considered complex interactions between listener groups, speaking styles, background noise, and SNRs. We predicted that (a) speaking style adaptations will improve word recognition less in SSN and 2T babble, where accuracy is expected to be relatively high as performance approaches ceiling, whereas the benefit is expected to be largest in 6T babble, where performance is expected to be lowest. (b) Compared with the babble maskers and CS, the benefit from NAS will be higher in SSN—in the matched target–masker condition—where the masker listeners heard is the same SSN used to elicit the production of NAS. (c) Nonnative listeners will benefit less from signal enhancements (speaking styles, more beneficial SNR) compared with native listeners. (d) Native English listeners will be more negatively affected by IM that consists of English babble speech compared with nonnative English listeners for whom the target signal and the babble are presented in their second language (L2).

Experiment 1: Intelligibility of CS and NAS in SSN and in Multitalker Babble for Native Listeners

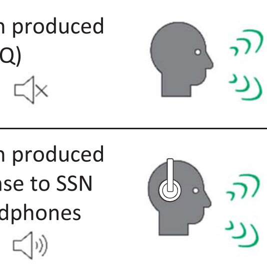
Method

Listeners

One hundred twenty listeners participated in the word-recognition-in-noise task. Participants were native English speakers (73 women; two participants declined to answer, and response from one participant was not recorded) with an average age of 20.3 years (*SD* = 4, range: 18–40). Our

Table 1. Overview of the speaking style elicitation conditions (left panel) and listening conditions (right panel).

Speaking style elicitation conditions		Listening conditions		
1. Speech produced in quiet (Q)	Conversational (CO) Clear speech (CS)			
2. Speech produced in response to SSN over headphones	Conversational (CO) Clear speech (CS)			
		2T babble masker	6T babble masker	SSN masker



Note. Conversational (CO) and clear speaking (CS) styles were elicited under two conditions: (a) in quiet (Q) and (b) in response to the speech-shaped noise (SSN) played over the headphones (noise-adapted speech). Listeners heard target sentences presented in two-talker (2T) babble, six-talker (6T) babble, or SSN and at two signal-to-noise ratios.

participants were considered “native” because they self-reported to have been born and raised in environments where English was the primary language. All listeners were students at The University of Texas at Austin (UT Austin). The study was approved by the institutional review board at UT Austin, and participants provided written consent to use their responses anonymously. Before beginning the experimental task, they filled out a detailed language background questionnaire adapted from the Language Experience and Proficiency Questionnaire (Marian et al., 2007). All listeners passed a hearing screening bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz. They received class credit or a small monetary compensation for participation.

Sentence Materials

The stimuli used in this study were a subset of 40 sentences (e.g., “Mice like to eat cheese”; Fallon et al., 2002) used in the work of Smiljanić and Gilbert (2017a, 2017b). Ten young adult talkers (18–29 years old, M_{age} = 21, five women), identified as native monolingual speakers of American English, read all sentences using CO and then CS. The conversational and clear sentences were first elicited in quiet (Q), and then, both styles were produced in response to noise (NAS) presented over headphones (80 dB SPL). The SSN that talkers heard over headphones had been generated by shaping white noise to match the long-term average spectrum of the 6T babble (three female and three male talkers; Van Engen & Bradlow, 2007). When eliciting CO in quiet, talkers were told to speak in a casual manner as if they were talking to a friend or a family member. For CS elicited in quiet, they were instructed to speak as if they were communicating with someone who has low proficiency in English and who has difficulty following them conversationally. When eliciting NAS, the talkers were told that they would hear background noise and to imagine they were speaking with someone in a noisy place. They followed the same instructions for producing CO and CS in quiet while hearing the SSN interference through headphones.

Forty unique meaningful sentences produced by each of the 10 talkers in four speaking styles, for a total of 1,600 sentences (40 sentences × 4 speaking styles × 10 talkers), were used in the word recognition experiments. All sentences were normalized for root-mean-square amplitude and mixed with SSN and competing speech. The competing speech consisted of either 2T (two female talkers) or 6T (three female and three male talkers) babble (Van Engen & Bradlow, 2007). Different portions of the multitalker babble were mixed with different target sentences to avoid repeating the same background speech in each sentence. We used two SNRs, “easier” (−5 dB) and “harder” (−7 dB), to get a range of intelligibility scores across speaking style enhancements and maskers. We chose these two SNR levels for two reasons. First, we were hoping to avoid the ceiling and floor performance in any listening condition. If the task was too hard (close to 0% accuracy) or too easy (close to 100% accuracy), we would not be able to observe CS and NAS intelligibility benefits or interactions with maskers. We hoped to get the accuracy range in CO produced in quiet and presented in the SSN masker such that we could observe performance improvement from speaking style enhancements in the harder (6T babble) listening condition and in the easier (2T babble and SSN) listening conditions. The second goal in choosing the SNRs in Experiment 1 (−5 and −7 dB) and Experiment 2 (−3 and −5 dB) was to attempt to equate the performance across the two listener groups in CO produced in quiet and presented in the SSN masker. This would ensure that group differences in performance relative to the baseline condition are not confounded by starting-level differences (Bradlow & Alexander, 2007; Sommers, 1996, 1997). The SNR levels were determined based on previous work using the same materials (Smiljanić & Gilbert, 2017b), through pilot data with a small number of participants using the materials from this study, and on previous work (Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011). Acoustic analyses for the sentences used in this study were reported in the work of Smiljanić and Gilbert (2017b).

Procedure

One hundred twenty listeners participated in one of the six listening conditions (3 maskers × 2 SNRs). Each listener thus heard target sentences mixed with SSN, 2T babble, or 6T babble at an easier (−5 dB) or a more difficult (−7 dB) SNR. Following the procedure used in the work of Smiljanić and Gilbert (2017b), the experiment began with five practice sentences mixed with the masker used in that listening condition (SSN or multitalker babble) to familiarize listeners with the task and the nature of the stimuli; practice sentences consisted of talkers and stimuli not included in the test stimuli. In the babble conditions, the participants were told that they will first hear the background speech and that the target speaker will be briefly delayed in each trial. During the test phase, listeners heard four sentences from each of the 10 speakers, for a total of 40 unique sentences. The four sentences included from each talker were produced in a different speaking style (e.g., Sentence 1 in CO produced in quiet, Sentence 12 in CS produced in quiet, Sentence 30 in CO produced in noise, and Sentence 39 in CS produced in noise, all from the same talker). Each listener heard a different combination of 40 unique sentences produced by all 10 talkers and across all four speaking styles. The 40 items were randomized for each listener, and the pairing of sentences to speaking styles was counterbalanced across listeners. This ensured that each listener heard every talker in every speaking style without sentence repetition. Speaking style was thus a within-subject factor, whereas masker type and SNR were across-subjects factors. The experiment was presented in MATLAB. Listeners were instructed to write what they heard, typing one sentence at a time on a keyboard after a stimulus presentation. They were asked to guess when they were unsure or unable to understand the complete sentence. After each trial, participants initiated the next trial by clicking a button. Items were presented only once. For each trial, the noise started 500 ms before the target sentence and stopped 500 ms after the target ended. The experimental session lasted approximately 30 min.

Analyses

All sentences contained three target words that were scored "1" (correct) or "0" (incorrect). Key words with omitted morphemes were considered incorrect responses, but obvious spelling errors or homophones were considered correct.

Data were analyzed with logistic mixed-effects regression, fitted using the glmer function in the lme4 package in R (Bates et al., 2015). To determine the model with the best fit, regressions were performed using a backward stepwise selection method with the drop1() function (Heumann et al., 2016, p. 282). All models had word recognition (correct vs. incorrect) as a dependent variable with a binomial link function and random intercepts for talkers and stimuli. The best model for the analysis of native English listeners' responses included speaking style (CO vs. CS, Q vs. NAS), masker type (SSN, 2T babble, 6T babble), and SNR (easier, harder) as fixed effects. Interactions between the two types of speaking style, between masker and each of the two speaking styles, and between masker and SNR were included.

Type III Wald chi-square tests were used to determine significance for model parameters. Significant results were further analyzed with estimated marginal means using the emmeans package in R (Lenth, 2018). Effect size is explained in terms of odds ratios (*OR*s). *OR*s provide probability estimates that indicate a change in the odds of the outcome occurring for a certain predictor. An *OR* greater than 1 means that a predictor was positively related to the outcome, whereas an *OR* smaller than 1 means a predictor was negatively related to the outcome.

Results

The proportion of key words correctly identified for each masker and SNR for native English listeners is shown in Figure 1. Results from Type III Wald chi-square tests as well as the 95% confidence intervals (CIs) and p values for *OR*s from post hoc comparisons are shown in Table 2.

Regression analyses revealed that both speaking style modifications were significant predictors of word recognition accuracy. CS was significantly more intelligible than CO, and NAS was significantly more intelligible than speech produced in quiet. The interaction between the two speaking style modifications, $\chi^2(1) = 28.95$, $p < .001$, showed that relative to CO, CS improved intelligibility when produced in quiet and when produced in response to noise (CS+NAS). CO and CS were more intelligible when produced in response to noise (CO+NAS and CS+NAS) than when produced in quiet. NAS was, overall, significantly more intelligible than CO and CS. Word recognition accuracy was highest when both modifications were combined (CS+NAS).

Speaking styles also interacted with masker, CS × Masker: $\chi^2(2) = 32.21$, $p < .001$; NAS × Masker: $\chi^2(2) = 19.06$, $p < .001$. In all three maskers, CS and NAS improved word recognition accuracy compared with CO and speech produced in quiet, respectively. Pairwise comparisons showed that the intelligibility benefit for both CS and NAS was highest in SSN. The odds of accurate word recognition for CS compared with that for CO were 2.54 times higher in SSN compared with 1.49 times higher in 6T babble and 1.78 times higher in 2T babble. For NAS compared with speech in quiet, the odds of accurate word recognition were 2.88 times higher in SSN compared with 2.25 times higher in 6T babble and 1.88 times higher in 2T babble.

A significant interaction between SNR and masker, $\chi^2(2) = 380.51$, $p < .001$, revealed that word recognition accuracy was significantly higher in 2T babble than in 6T babble as well as in SSN at both SNRs. While there was no significant difference between 6T babble and SSN at an easier SNR ($p = .54$), intelligibility in SSN was significantly better than that in 6T babble at a harder SNR. Finally, word recognition accuracy was significantly higher at an easier SNR compared with that at a harder SNR for both babble noise conditions. The odds of successful key word identification increased by 394% in 6T babble and by 202% in 2T babble. Unexpectedly, word recognition accuracy was significantly lower ($p = .02$) at an easier SNR compared with that at a more difficult SNR, in SSN only. As we do not currently have an explanation, it would be important to try and replicate this finding in the future to eliminate the possibility of a spurious result.

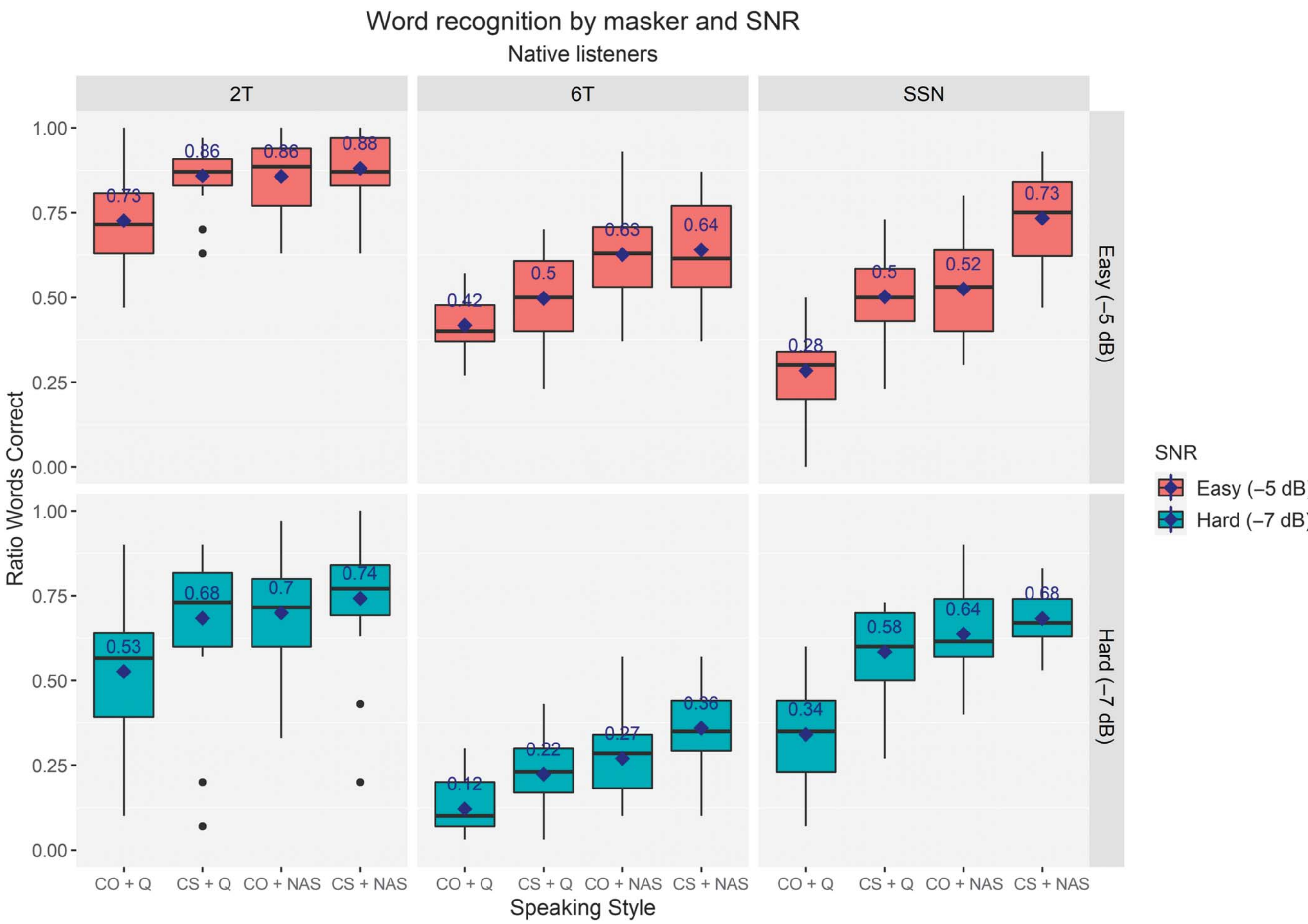
Experiment 2: Intelligibility of CS and NAS in SSN and in Multitalker Babble for Nonnative Listeners

Method

Listeners

One hundred twenty listeners participated in the word-recognition-in-noise task (71 women) with an

Figure 1. Native English listeners' key word correct score for four different speaking styles, namely, conversational speech produced in quiet (CO+Q), clear speech produced in quiet (CS+Q), noise-adapted conversational speech (CO+NAS), and noise-adapted clear speech (CS +NAS), presented in two-talker (2T) babble (left panels), six-talker (6T) babble (middle panels), and speech-shaped noise (SSN; right panels). The top panels show performance at an easier signal-to-noise ratio (SNR; −5 dB), and the bottom panels show performance at a harder SNR (−7 dB). Error bars are standard errors of the mean. Diamond symbols and numerical values (in blue) indicate mean word recognition accuracy for each speaking style and masker.



average age of 22.8 years (*SD* = 4.6, range: 18–38). On average, they were first exposed to English after the age of 8.5 years (range: 2–19). Participants' language background information is provided in Table 3. Participants were recruited at UT Austin. The study was approved by the institutional review board at UT Austin, and participants provided written consent to use their responses anonymously. All listeners passed a hearing screening bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz. They received class credit or a small monetary compensation for participation.

Sentence Materials

The stimuli in Experiment 2 were the same 40 sentences produced in CO and CS in quiet and in response to noise (NAS) by 10 talkers as in Experiment 1. The only difference between the two experiments, besides the listener groups, is in the SNR levels used in the listening task. For nonnative listeners, we adjusted the SNR to −3 dB (easier) and −5 dB (harder), as opposed to −5 and −7 dB SNRs, which were used for native listeners. Because of the well-established disproportionate effect of noise on nonnative speech recognition in noise, we used a slightly more beneficial SNR for nonnative than native listeners (Bradlow & Alexander, 2007; Mayo et al., 1997; Smiljanić & Bradlow, 2011; Van Summers et al., 1988). The goal was to assess the CS and NAS intelligibility benefits relative to a similar baseline performance (CO produced in quiet and presented in SSN) for the two listener groups (see a similar approach in Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011; Sommers, 1997). Any observed group differences in word

Table 2. Results of the linear mixed-effects logistic regression for native English listeners' intelligibility.

Fixed effects	χ^2	*df*	Pr(> χ^2)
Q vs. NAS	160.82	1	< .001
CO vs. CS	5.48	1	.0193
Masker (SSN, 2T, 6T)	315.43	2	< .001
SNR (–5 dB, –7 dB)	520.57	1	< .001
NAS × CS	28.95	1	< .001
	OR	**95% CI**	***p***
CO+NAS vs. CO+Q	2.86	[2.46, 3.32]	< .001
CS+NAS vs. CS+Q	1.85	[1.59, 2.16]	< .001
CS+Q vs. CO+Q	2.34	[2.02, 2.72]	< .001
CS+NAS vs. CO+NAS	1.52	[1.31, 1.77]	< .001
CO+NAS vs. CS+Q	1.22	[1.05, 1.41]	< .001
	χ^2	*df*	Pr(> χ^2)
CS × Masker	32.21	2	< .001
	OR	**95% CI**	***p***
CS vs. CO (presented in SSN masker)	2.54	[2.09, 3.09]	< .001
CS vs. CO (presented in 6T masker)	1.49	[1.22, 1.82]	< .001
CS vs. CO (presented in 2T masker)	1.78	[1.43, 2.22]	< .001
	χ^2	*df*	Pr(> χ^2)
NAS × Masker	19.06	2	< .001
	OR	**95% CI**	***p***
NAS vs. Q (presented in SSN masker)	2.88	[2.37, 2.50]	< .001
NAS vs. Q (presented in 6T masker)	2.25	[1.84, 2.75]	< .001
NAS vs. Q (presented in 2T masker)	1.88	[1.51, 2.33]	< .001
	χ^2	*df*	Pr(> χ^2)
Masker × SNR	380.51	2	< .001
	OR	**95% CI**	***p***
2T vs. SSN (–5 dB)	6.11	[4.87, 7.65]	< .001
2T vs. 6T (–5 dB)	5.33	[4.26, 6.66]	< .001
SSN vs. 6T (–5 dB)	1.15	[0.95, 1.39]	.54
2T vs. SSN (–7 dB)	1.63	[1.34, 1.98]	< .001
2T vs. 6T (–7 dB)	8.72	[7.06, 10.77]	< .001
SSN vs. 6T (–7 dB)	5.36	[4.35, 6.60]	< .001
–5 dB vs. –7 dB (presented in 6T masker)	4.94	[4.03, 6.07]	< .001
–5 dB vs. –7 dB (presented in 2T masker)	3.02	[2.41, 3.78]	< .001
–5 dB vs. –7 dB (presented in SSN masker)	0.81	[0.66, 0.98]	.016

Note. Estimated odds ratios and *z* statistics from post hoc comparisons are also provided for interaction terms. Q = speech produced in quiet; NAS = noise-adapted speech; CO = conversational speaking style; CS = clear speaking style; SSN = speech-shaped noise; 2T = two-talker babble; 6T = six-talker babble; SNR = signal-to-noise ratio; *OR* = odds ratio; CI = confidence interval.

recognition in more difficult or easier listening conditions would thus not be confounded by the starting/baseline-level differences. As in Experiment 1, it was also important to avoid ceiling or floor performance in any listening conditions, which would obscure any CS and NAS intelligibility benefits. The SNR levels were determined based on previous work using the same materials (Smiljanić & Gilbert, 2017b), through pilot data with a small number of participants using the materials from this study, and on previous work (e.g., Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011). Note that the same −5 dB SNR is considered the "harder" condition for nonnative listeners and the "easier" condition for native listeners, which we will return to below.

Procedure

The experimental procedure was identical to that for Experiment 1.

Analyses

The results were scored and analyzed as in Experiment 1. The best model for the analysis of nonnative

Table 3. Language background information for nonnative listeners (N = 120).

	M	*SD*	Range
Age of first exposure to English (in years)	8.5	3.1	2–19
Age of arrival to USA (in years)	16.0	7.9	0–35
Time spent in USA (in years)	6.8	6.1	0.08–30
Daily exposure to English[a]	4.7	0.6	2–5
Daily exposure to L1	4.5	0.9	2–5
Self-estimated proficiency in English[b]	3.9	1.0	1–5
Self-estimated proficiency in L1	4.3	0.9	1–5
L1	Spanish (n = 44), Mandarin (n = 33), Korean (n = 12), French (n = 4), Vietnamese (n = 3), Hindi (n = 3), Arabic (n = 3), Portuguese (n = 2), Russian (n = 2), Cantonese (n = 2), Turkish (n = 2), German (n = 1), Japanese (n = 1), Czech (n = 1), Hebrew (n = 1), Farsi (n = 1), Dutch (n = 1), Nepali (n = 1), Yoruba (n = 1), Polish (n = 1), Amharic (n = 1)		

Note. The following information in this table was not provided by some participants: age of first exposure to English (n = 2), daily exposure to English/L1 (n = 31), self-estimated proficiency in English (n = 3), self-estimated proficiency in L1 (n = 7). L1 = first language.

[a]For each language, self-estimated amount of daily exposure on a scale from 1 (*no current exposure*) to 5 (*constant exposure*). [b]For each language, average of self-estimated proficiency for each skill, that is, writing, speaking, reading, and listening, on a scale from 1 (*low*) to 5 (*high*). Ratings by 28 participants were converted to a 5-point scale from a 10-point scale used in a previous version of a language questionnaire used in our lab.

English listeners' accuracy data was the same as for the native listeners. It included speaking style (CO vs. CS, Q vs. NAS), masker type (SSN, 2T babble, 6T babble), and SNR (easier, harder) as fixed effects. Interactions between the two types of speaking style, between masker type and each of the two speaking styles, and between masker and SNR were considered.

Results

The proportion of words correctly identified for each masker and SNR for nonnative listeners is shown in Figure 2. Results from Type III Wald chi-square tests as well as the 95% CIs and p values for *OR*s from post hoc comparisons are shown in Table 4.

As was found for native listeners, regression analyses showed that CS, $\chi^2(1) = 10.44$, $p < .01$, and NAS, $\chi^2(1) = 78.42$, $p < .001$, were significant predictors of key word identification for nonnative listeners. The two-way CS × NAS interaction was also significant, $\chi^2(1) = 32.93$, $p < .001$. Pairwise comparisons showed that speaking clearly and in response to noise significantly improved word recognition compared with CO and speech produced in quiet, respectively. When both speaking style modifications were produced simultaneously (CS+NAS), intelligibility was significantly greater than when produced alone. In contrast to native listeners, nonnative listeners benefited equally from NAS and CS in word recognition in noise.

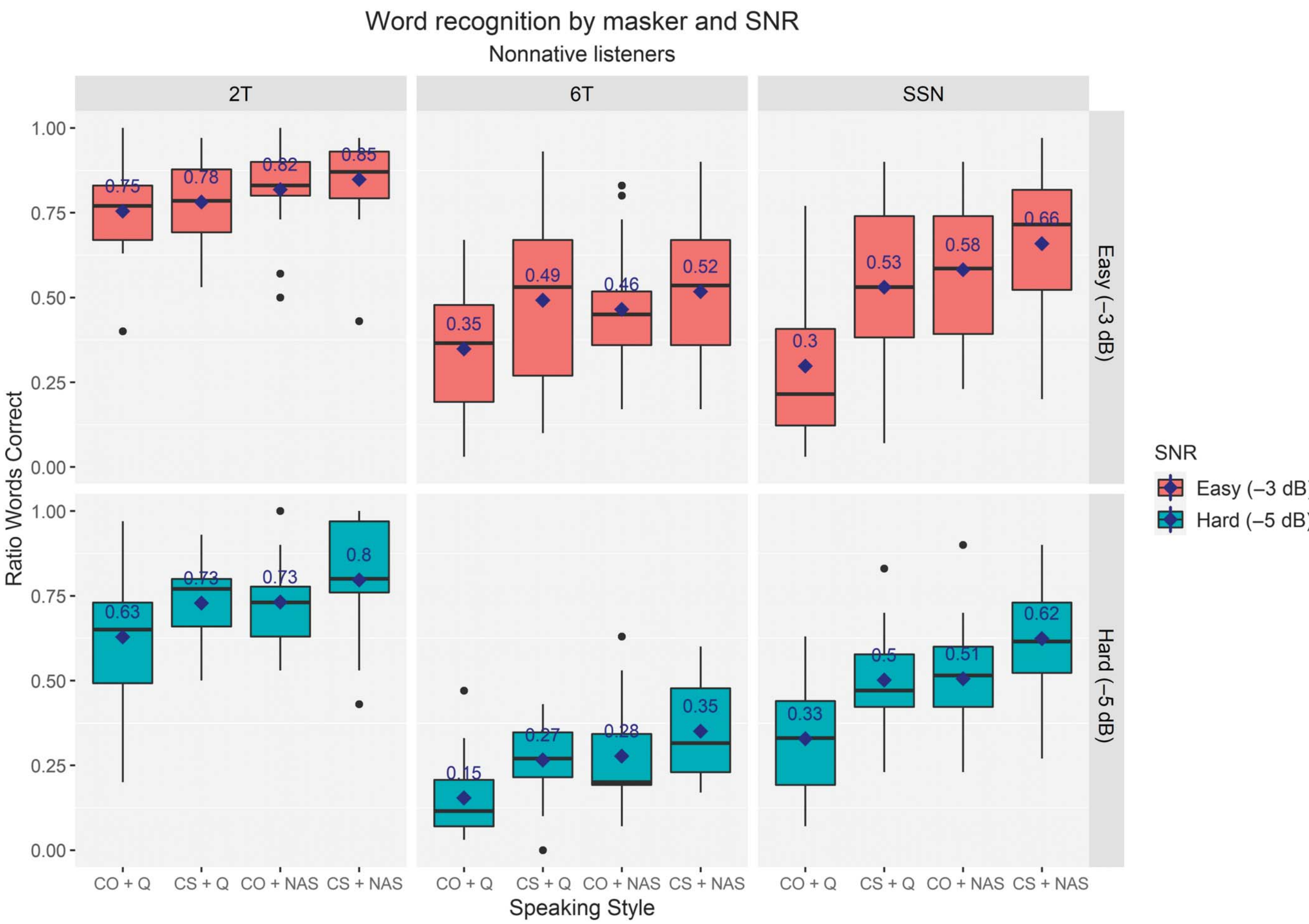
The analyses further revealed significant two-way interactions between speaking style and masker, CS × Masker: $\chi^2(2) = 13.84$, $p < .001$; NAS × Masker: $\chi^2(2) = 30.77$, $p < .001$. Speaking clearly and in response to noise, separately and combined, significantly improved intelligibility compared with CO and speech produced in quiet in all three maskers, but the magnitude of the effect varied. Compared with CO, the odds of correct word recognition for CS more than doubled in SSN and increased by 1.61 in 6T babble and by 1.47 in 2T babble. Compared with speech produced in quiet, the odds of correct word recognition for NAS increased by 2.46 in SSN, by 1.61 in 6T babble, and by 1.53 in 2T babble.

Lastly, an interaction between SNR and masker was also significant, $\chi^2(2) = 68.16$, $p < .001$. Contrast analyses revealed that, at both SNRs, word recognition accuracy was highest in 2T babble, followed by SSN, and lowest in 6T babble. Finally, the odds of successful key word identification were significantly higher for the easier than for the more difficult SNR in all three maskers. The odds increased by 171% in 6T babble and by 27% in SSN.

Listener Group Comparison: Native Versus Nonnative English Listeners

The above analyses assessed the effect of speaking style, masker, and SNR on word recognition accuracy for native (Experiment 1) and nonnative (Experiment 2) listeners separately. To shed light on the listener-related differences more directly, here, we compare word recognition accuracy for native and nonnative listeners in the overlapping −5 dB SNR condition. In order to visualize the results with respect to how much each speaking style, separately and in combination, enhanced intelligibility for native and nonnative listeners in different maskers, the raw data are presented in Figure 3 as average improvement scores relative to the baseline intelligibility in the most

Figure 2. Nonnative English listeners' key word correct score for four different speaking styles, namely, conversational speech produced in quiet (CO+Q), clear speech produced in quiet (CS+Q), noise-adapted conversational speech (CO+NAS), and noise-adapted clear speech (CS+NAS), presented in the presence of two-talker (2T) babble (left panels), six-talker (6T) babble (middle panels), and speech-shaped noise (SSN; right panels). The top panels show performance at an easier signal-to-noise ratio (SNR; −5 dB), and the bottom panels show performance at a harder SNR (−7 dB). Error bars are standard errors of the mean. Diamond symbols and numerical values (in blue) indicate mean word recognition accuracy for each speaking style and masker.



difficult condition, that is, CO produced in quiet (e.g., intelligibility of CO produced in response to noise [CO+NAS] minus intelligibility of CO produced in quiet [CO in quiet] divided by intelligibility of CO in quiet).

Analysis

The best model for the group comparison included speaking style (CO vs. CS, Q vs. NAS), masker type (SSN, 2T babble, 6T babble), and listener group (native, nonnative) as fixed effects. The model also included two-way interactions for the listener group and each fixed effect as well as three-way interactions between the listener group and the two speaking styles and between the listener group, each speaking style, and masker type. The four-way interaction was also examined, but it was not significant and, thus, will not be further discussed.

Results

Combined word recognition accuracy for the two listener groups in the −5 dB SNR condition is shown in Figure 4. Results from Type III Wald chi-square tests as well as the 95% CIs for each *OR* and *p* values are provided in Table 5.

Results showed significant interactions between listener group and speaking styles, $\chi^2(2) = 23.86$, $p < .001$. Three-way interactions were also significant, Listener Group × Masker × NAS: $\chi^2(4) = 45.16$, $p < .001$; Listener Group × Masker × CS: $\chi^2(4) = 20.08$, $p < .001$. Post hoc analyses showed that native listeners benefited significantly more than nonnative listeners did from CS and NAS adaptations. However, the significant interactions revealed that this varied across the three maskers: Compared with nonnative listeners, native listeners' odds of successful word recognition for CS

Table 4. Results of the linear mixed-effects logistic regression for nonnative English listeners' intelligibility.

Fixed effects	χ^2	*df*	Pr(> χ^2)
Q vs. NAS	78.42	1	< .001
CO vs. CS	10.44	1	< .01
Masker (SSN, 2T, 6T)	365.16	2	< .001
SNR (–3 dB, –5 dB)	219.38	1	< .001
NAS × CS	32.93	1	< .001
	OR	**95% CI**	***p***
CO+NAS vs. CO+Q	2.28	[1.97, 2.65]	< .001
CS+NAS vs. CS+Q	1.45	[1.26, 1.68]	< .001
CS+Q vs. CO+Q	2.13	[1.84, 2.46]	< .001
CS+NAS vs. CO+NAS	1.35	[1.17, 1.57]	< .001
CO+NAS vs. CS+Q	1.08	[0.93, 1.24]	1.00
	χ^2	*df*	Pr(> χ^2)
CS × Masker	13.84	2	< .001
	OR	**95% CI**	***p***
CS vs. CO (presented in SSN masker)	2.08	[1.72, 2.49]	< .001
CS vs. CO (presented in 6T masker)	1.61	[1.32, 1.96]	< .001
CS vs. CO (presented in 2T masker)	1.47	[1.18, 1.83]	< .001
	χ^2	*df*	Pr(> χ^2)
NAS × Masker	30.77	2	< .001
	OR	**95% CI**	***p***
NAS vs. Q (presented in SSN masker)	2.46	[2.04, 2.98]	< .001
NAS vs. Q (presented in 6T masker)	1.61	[1.32, 1.96]	< .001
NAS vs. Q (presented in 2T masker)	1.53	[1.23, 1.89]	< .001
	χ^2	*df*	Pr(> χ^2)
Masker × SNR	68.16	2	< .001
	OR	**95% CI**	***p***
2T vs. SSN (–5 dB)	4.63	[3.75, 5.72]	< .001
2T vs. 6T (–5 dB)	6.31	[5.11, 7.79]	< .001
SSN vs. 6T (–5 dB)	1.36	[1.13, 1.65]	< .001
2T vs. SSN (–7 dB)	3.52	[2.89, 4.29]	< .001
2T vs. 6T (–7 dB)	10.28	[8.33, 12.69]	< .001
SSN vs. 6T (–7 dB)	2.92	[2.39, 3.56]	< .001
–3 dB vs. –5 dB (presented in 6T masker)	2.71	[2.23, 3.30]	< .001
–3 dB vs. –5 dB (presented in 2T masker)	1.66	[1.34, 2.06]	< .001
–3 dB vs. –5 dB (presented in SSN masker)	1.27	[1.05, 1.53]	.004

Note. Estimated odds ratios and *z* statistics from post hoc comparisons are also provided for interaction terms. Q = speech produced in quiet; NAS = noise-adapted speech; CO = conversational speaking style; CS = clear speaking style; SSN = speech-shaped noise; 2T = two-talker babble; 6T = six-talker babble; SNR = signal-to-noise ratio; *OR* = odds ratio; CI = confidence interval.

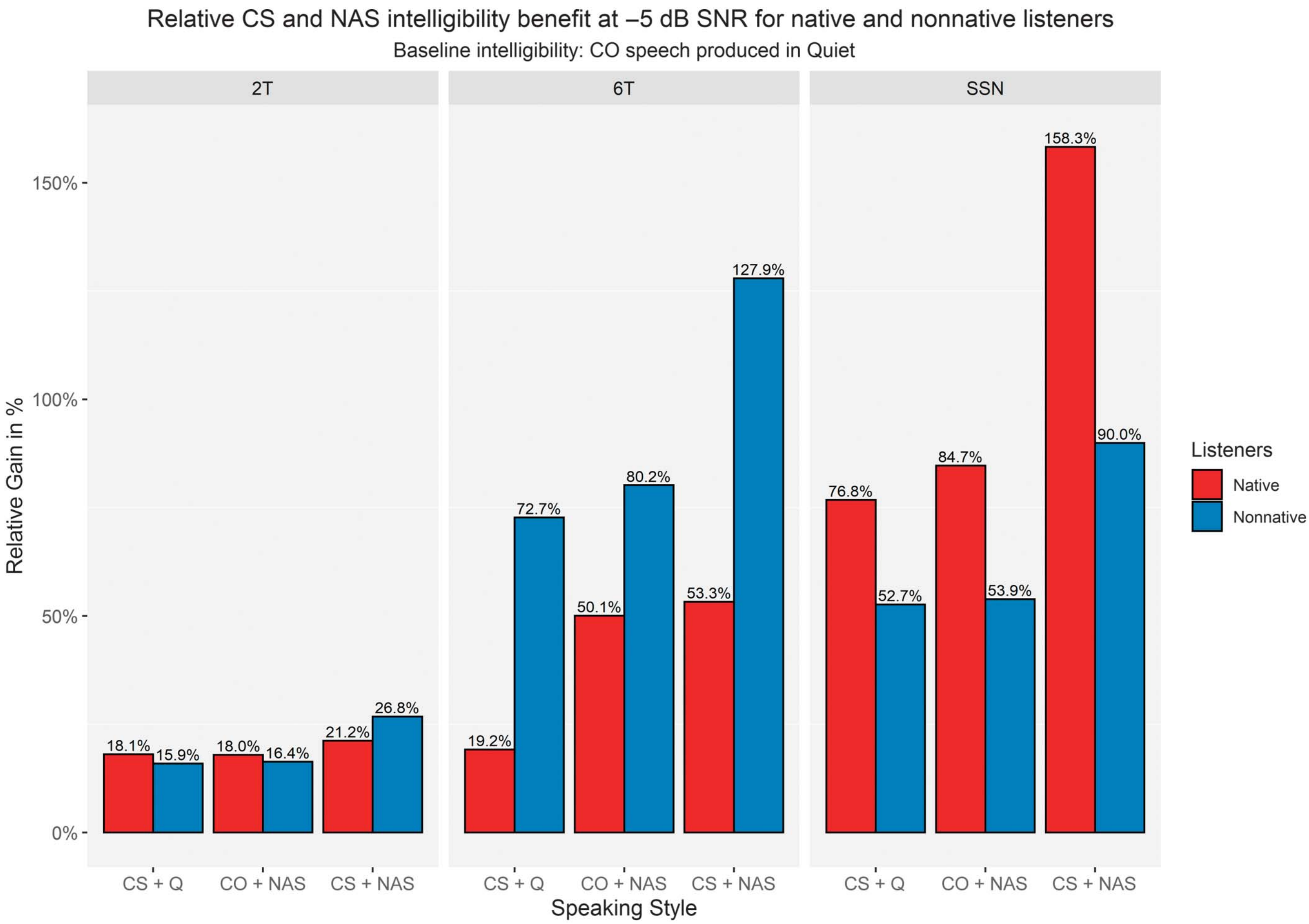
style were 121% higher in 2T babble, 278% higher in 6T babble, and 60% higher in SSN. For sentences produced in NAS, native listeners' odds of successful word identification were 140% higher in 2T babble, 349% higher in 6T babble, and 50% higher in SSN compared with nonnative listeners.

As predicted, the similarly small proportional gain for any speaking style enhancement for both listener groups was found in 2T babble, where intelligibility was high overall. Larger proportional gains were found in 6T babble and SSN for both groups. Compared with nonnative listeners, native listeners benefited significantly more from CS produced in quiet, CO+NAS, and CS+NAS when listening to target speech masked with SSN. In the 6T babble masker, however, nonnative listeners benefited significantly more from the speaking style enhancements even though their performance was lower overall.

Discussion

The goal of this study was to compare the intelligibility benefit associated with two speaking style modifications produced in response to listener perceptual difficulty (CS)

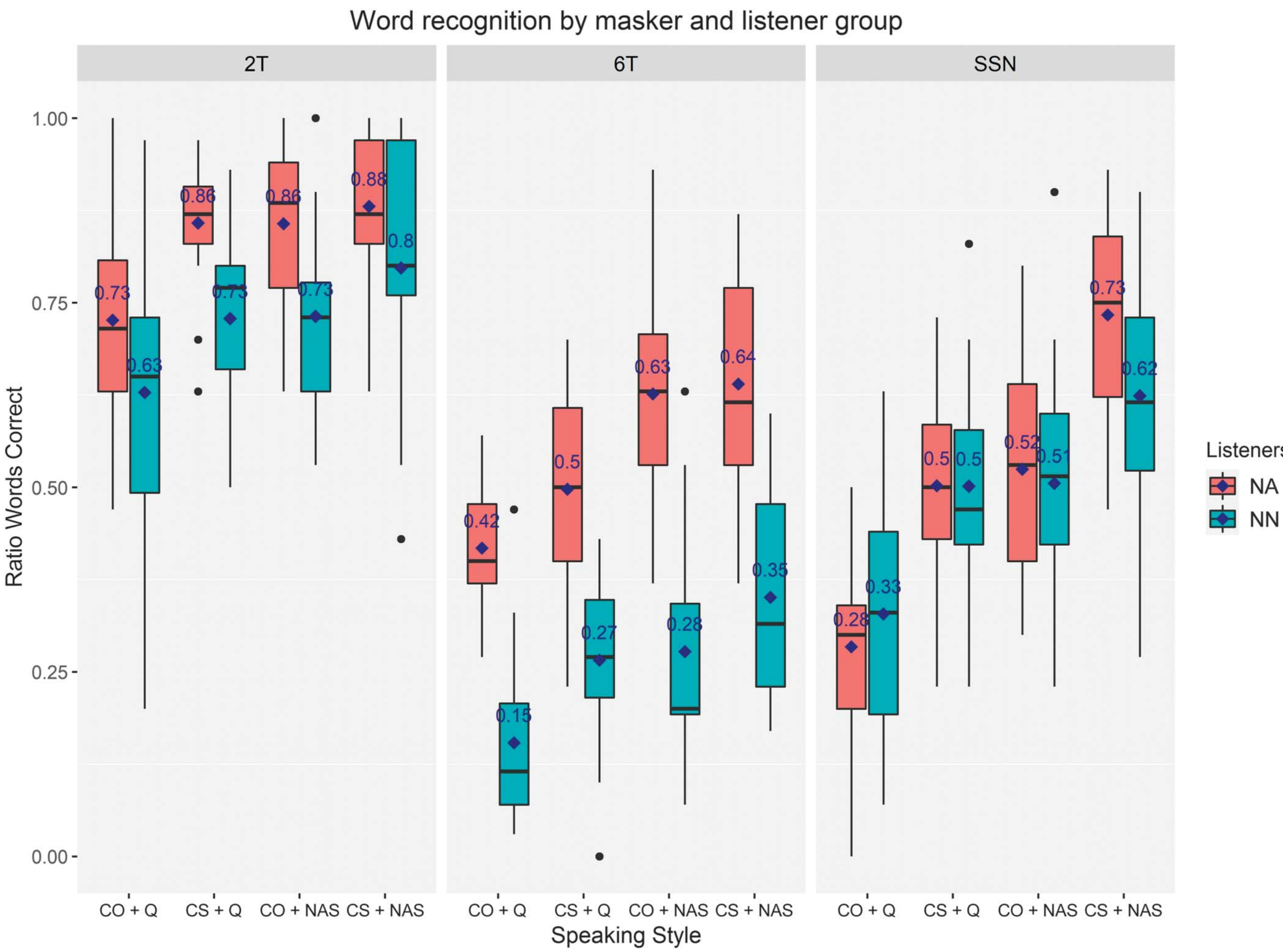
Figure 3. Native (red bars) and nonnative (blue bars) listeners' intelligibility benefits for three speaking styles, namely, clear speech produced in quiet (CS+Q), noise-adapted conversational speech (CO+NAS), and noise-adapted clear speech (CS+NAS), relative to intelligibility in baseline conversational speech (CO) produced in quiet. Intelligibility gains vary for sentences presented in the presence of two-talker (2T) babble (left panels), six-talker (6T) babble (middle panels), and speech-shaped noise (SSN; right panels) at −5 dB SNR.



and environmental noise (NAS), separately and in combination. We focused on the effectiveness of CS and NAS for improving word recognition in the presence of maskers varying in the amount of EM and IM contributions and at different SNRs. We also assessed whether native and nonnative listeners derived similar intelligibility benefit from CS and NAS under different listening conditions. Consistent with our predictions, the main effects of the tested factors were in line with previous work: (a) CS and NAS, separately and in combination, were significantly more intelligible than CO and speech produced in quiet, for both native and nonnative listeners; (b) intelligibility of the target speech presented in 2T babble was better than that in 6T babble and SSN; and (c) speech was more intelligible at an easier SNR than at a harder SNR. The findings revealed several significant interactions demonstrating that the CS and NAS intelligibility benefits were determined by the listening environment, both the masker and the SNR. We will discuss these results in turn next.

Each speaking style modification (CS and NAS) alone improved intelligibility, and their combination (CS+NAS) resulted in the highest intelligibility in the presence of all three maskers (2T babble, 6T babble, and SSN) and at both SNR levels ("easier" and "harder"). This finding is in line with previous work that demonstrated improved intelligibility of CS and NAS over CO and speech produced in quiet separately (Cooke et al., 2014; Godoy et al., 2013; Pichora-Fuller et al., 2010; Smiljanić, 2021). The results expand on previous findings by showing the CS advantage in maskers other than SSN (Bradlow & Alexander, 2007; Liu & Zeng, 2006; Van Engen, 2017; but see Van Engen et al., 2014, for the CS benefit in SSN and multitalker maskers). Furthermore, the findings provide new evidence that intelligibility benefit is largest when the two speaking style enhancements are combined, CS+NAS, in multitalker babble

Figure 4. Native (NA, red bars) and nonnative (NN, blue bars) English listeners' key word correct score for four speaking styles at −5 dB SNR, namely, conversational speech produced in quiet (CO+Q), clear speech produced in quiet (CS+Q), noise-adapted conversational speech (CO+NAS), and noise-adapted clear speech (CS+NAS), presented in two-talker (2T) babble (left panel), six-talker (6T) babble (middle panel), and speech-shaped noise (SSN; right panel). Error bars are standard errors of the mean. Diamond symbols and numerical values (in blue) indicate mean word recognition accuracy for each speaking style and masker.



maskers (cf. Gilbert et al., 2014; Smiljanić & Gilbert, 2017b). As in previous work, both native and nonnative listeners benefited from each intelligibility-enhancing speaking style (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Cooke & Garcia Lecumberri, 2012; Cooke et al., 2008; Smiljanić & Bradlow, 2011). Here, we extended previous findings to show that, similar to native listeners, nonnative listeners benefited most from the combined CS+NAS modifications. When talkers increased vocal effort in response to noise in combination with deliberate, clear productions, listeners' word recognition accuracy improved significantly more than when each modification was produced separately. Acoustic–phonetic modifications in the form of speaking style adjustments are efficient strategies for improving speech perception in noise for both native and nonnative listeners in a variety of listening conditions.

With regard to maskers, listeners were least affected by a predominantly informational masker (2T babble) and most affected by 6T babble with the combined IM and EM. Word recognition was intermediate for the purely energetic masker (SSN; Brouwer et al., 2012; Calandruccio et al., 2010; Van Engen et al., 2014). The word recognition accuracy difference among the maskers is in line with the glimpsing model of speech perception whereby 2T babble exerts less EM and provides more and larger spectrotemporal dips compared with 6T babble and SSN. For energetically more constant maskers, such as SSN or 6T babble, temporal glimpses are rarer, hindering access to the target speech and, hence, resulting in worse performance. The available glimpses in 2T babble contain important information to support speech recognition in noise despite the lexical intrusion of the intelligible background speech. Note that in this study,

Table 5. Results of the linear mixed-effects logistic regression for native versus nonnative English listeners' intelligibility at −5 dB SNR.

Fixed effects	χ^2	*df*	Pr(> χ^2)
Q vs. NAS	79.12	1	< .001
CO vs. CS	0.73	1	.394
Masker (SSN, 2T, 6T)	198.66	2	< .001
Listener group (NA, NN)	170.66	1	< .001
CS × Listener Group	2.22	1	.137
NAS × Listener Group	0.0003	1	.986
Masker × Listener Group	83.01	2	< .001
NAS × CS × Listener Group	23.86	2	< .001
	OR	**95% CI**	***p***
NA vs. NN (in CS+Q)	2.08	[1.78, 2.44]	< .001
NA vs. NN (in CO+NAS)	2.37	[2.02, 2.77]	< .001
NA vs. NN (in CO+Q)	2.04	[1.74, 2.39]	< .001
NA vs. NN (in CS+NAS)	2.70	[2.29, 3.19]	< .001
	χ^2	*df*	Pr(> χ^2)
CS × Masker × Listener Group	45.16	4	< .001
	OR	**95% CI**	***p***
NA vs. NN (CS presented in 2T masker)	2.21	[1.76, 2.78]	< .001
NA vs. NN (CS presented in 6T masker)	3.78	[3.14, 4.54]	< .001
NA vs. NN (CS presented in SSN masker)	1.60	[1.34, 1.92]	< .001
	χ^2	*df*	Pr(> χ^2)
NAS × Masker × Listener Group	20.08	4	< .001
	OR	**95% CI**	***p***
NA vs. NN (NAS presented in 2T masker)	2.40	[1.91, 3.02]	< .001
NA vs. NN (NAS presented in 6T masker)	4.49	[3.74, 5.40]	< .001
NA vs. NN (NAS presented in SSN masker)	1.50	[1.25, 1.80]	< .001

Note. Estimated odds ratios and *z* statistics from post hoc comparisons are also provided for interaction terms. Q = speech produced in quiet; NAS = noise-adapted speech; CO = conversational speaking style; CS = clear speaking style; SSN = speech-shaped noise; 2T = two-talker babble; 6T = six-talker babble; NA = native listeners; NN = nonnative listeners; *OR* = odds ratio; CI = confidence interval.

we used a two-female masker in the 2T babble listening condition. It is likely that differences in voice characteristics between talkers of different genders lead to greater masking release intelligibility for the target male talkers than for the target female talkers (Brungart, 2001; Brungart et al., 2001; Ericson et al., 2004). Nonetheless, we expect that this masking release advantage was consistent such that the overall accuracy patterns would still hold across the speaking style conditions. In other words, while the overall intelligibility for the target female talkers could be somewhat lower in 2T babble than that for the target male talkers, the pattern of CS and NAS intelligibility benefits would be similar for all target talkers. Future studies examining speaking style intelligibility benefits should include listening conditions with mixed-gender and opposite-gender babble maskers.

A greater decrease in performance observed for 6T babble compared with SSN could be due to several reasons, including larger spectral variations and smaller modulation depths present in the multitalker babble compared with the highly modulated masker such as SSN (Simpson & Cooke, 2005) as well as from IM based on target–masker linguistic misallocation. The exact contribution of each of these sources should be investigated more in the future. The interference ranking across masker types seems to hold when the auditory and cognitive systems are performing a more difficult task. Calandruccio et al. (2010), for example, found that native English listeners' performance was better when listening to sentences presented with an English 2T babble masker than an SSN masker but only at a more difficult SNR (at −5 dB SNR compared with −3 dB SNR). Our results showed a superior performance in 2T babble but no difference between SSN and 6T babble at −5 dB SNR for native listeners. With increased difficulty, at −7 dB SNR, native listeners' performance was higher in 2T babble than that in both SSN and 6T babble, and performance in SSN was better than that in 6T babble. Thus, the difference between the word recognition accuracy in 6T babble and that in SSN became evident only in the "harder" SNR condition. In contrast, the difference among the three types of maskers already emerged at an easier −3 dB SNR for nonnative listeners who are performing a cognitively more demanding task of L2 listening in noise (Garcia Lecumberri et al., 2010). This shows that the difficulty of perceiving speech in noise is not just a monotonic function of SNR but is determined by an interaction with a masker type and listener characteristics.

Next, we turn to the interactions among the speaking styles and maskers. As mentioned above, CS and NAS led to improved speech perception in all three maskers for both listener groups. However, the benefit varied across the maskers. The intelligibility benefit was largest in SSN and smallest in 2T babble, with the intermediate benefit in 6T babble. The listening task was harder overall in SSN and 6T babble than in 2T babble, as evidenced by the lower accuracy scores. In these challenging listening conditions, CS and NAS modifications made larger contributions toward increased intelligibility. The larger NAS benefit in SSN than in 6T babble could be further attributed to the matched target–masker (the same SSN was used when producing NAS and as a masker in perception). The listeners could have benefited from the acoustic–phonetic modifications talkers produced specifically in response to SSN, giving them an advantage compared with a 6T babble masker that was not used to elicit NAS in the production task. This, however, cannot account for the CS advantage in the SSN masker compared with the 2T babble and 6T babble maskers. CS was produced in quiet and then mixed with the SSN masker for the listening test. As suggested above, larger spectral variations and smaller modulation depths in 6T babble than in SSN could contribute to it being the most challenging for listeners (Van Engen et al., 2014; Zinszer et al., 2019).

The relatively small intelligibility benefit of CS and NAS in 2T babble compared with that in SSN could be explained by an overall high accuracy for the target speech presented in this masker. Listeners' overall performance in 2T babble was significantly higher compared with that in the other two maskers at both SNR levels. Listeners likely reached their maximum performance in this condition with little room for improvement from speaking style adaptations. Even though we chose the particular SNR levels in an attempt to avoid ceiling performance and even though the two listener groups performed similarly in the baseline condition (CO produced in quiet and presented in SSN), the large intelligibility variation arising from the two speaking styles, masker types, different SNRs, and individual listener abilities in speech-in-noise perception, we failed to achieve this goal for speech presented in 2T babble, the easiest listening condition. Nonetheless, the results suggest that the speaking style enhancements could significantly improve intelligibility in the presence of maskers that allow substantial glimpsing opportunities when presented in more challenging listening conditions (e.g., increased noise levels). This result further highlights the need to tease apart the contributions of the various factors shaping intelligibility variation in daily communication.

Finally, turning to the listener group differences, the current results revealed that native listeners benefited more from NAS adaptations than did nonnative listeners. Smiljanić and Gilbert (2017a) showed that both speaking style modifications included a decrease in speaking rate and an increase in 1- to 3-kHz energy, sound pressure level, and vowel space area, but they also differed. For example, increased f_o mean and decreased jitter and shimmer were produced only in NAS, whereas increased frequency and duration of pauses were produced only in CS. The difference in the CS and NAS intelligibility benefits for the two listener groups suggests a more efficient use of NAS enhancements by native listeners to overcome the interference from maskers. It is, however, not the case that nonnative listeners cannot use NAS acoustic–phonetic modifications to aid their word recognition. For both listener groups, intelligibility of CO produced in response to noise is substantially improved over that of CO produced in quiet. The results instead suggest that nonnative listeners may be less efficient in utilizing some of the modifications. It is possible that the intelligibility benefit for native listeners was also enhanced because the same noise (SSN) that was used to elicit the production of NAS was also used as one of the maskers in the listening task. In contrast, CS production was elicited in quiet and was subsequently mixed with the SSN masker for the listening task. This suggests that acoustic–phonetic modifications produced in response to noise (NAS) place energy in the parts of the spectrum that provide release from EM during perception (cf. Cooke & Garcia Lecumberri, 2012). Nonnative listeners did not show the additional benefit from the matched target–masker condition. It remains to be determined to what extent this discrepancy in NAS benefit between native and nonnative listeners is a result of their ability to rely on the various acoustic–phonetic enhancements in guiding their word recognition and at what specific level of perceptual processing this difference arises.

A direct comparison of native and nonnative listeners' performance in the overlapping SNR condition showed that native listeners outperformed nonnative listeners in almost all listening conditions. This is in line with previous work demonstrating that nonnative listeners reliably perform worse than native listeners in both energetic and informational maskers (Cooke et al., 2008; Mayo et al., 1997; Nábělek & Donahue, 1984) even when proficiency in the target language is high, as was the case in our study (Brouwer et al., 2012; Van Wijngaarden et al., 2002). Nonnative and bilingual listeners have a qualitatively different target language model compared with native monolingual listeners (see Garcia Lecumberri et al., 2010, for a review). When perceiving speech in noise, listeners need to integrate bottom-up acoustic information with top-down linguistic knowledge (Rönnberg et al., 2013). This allows them to fill in missing information in the acoustically degraded speech signals by using linguistically driven predictions (Cooke, 2006). Nonnative listeners, even the highly proficient ones, due to a different experience with the target language, may be less able to draw on linguistic knowledge at all levels to alleviate the noise interference (Zinszer et al., 2019). Higher accuracy for native listeners was found even for the target speech presented in 2T babble, where we predicted that the intelligible masker is going to interfere more with the native listeners' word recognition due to the interference from the intelligible background speech (Garcia Lecumberri & Cooke, 2006; Van Engen, 2010; Van Engen & Bradlow, 2007). It is not clear if this effect is due to our nonnative listener group's high proficiency with the target language such that the misattribution of linguistic information from speech masker noise components to the target speech (and vice versa) was similar between the two listener groups or due to a combination of factors suggested above. More work with listeners who vary in target language proficiency is needed to shed light on these questions.

Importantly, we showed that nonnative listeners significantly benefited from CS and NAS modifications, separately and in combination, improving their speech perception in different maskers and noise levels. In contrast to some previous work and to our predictions, we did not, however, find that native listeners benefited more from CS and NAS modifications in all listening conditions. Smaller CS and NAS intelligibility increases compared with native listeners found in the presence of the SSN masker align with previous work (Bradlow & Bent, 2002; Cooke et al., 2008; Smiljanić & Bradlow, 2011) and are consistent with the difficulty of performing the task in L2. Bradlow and Alexander (2007), for example, demonstrated that native listeners' word recognition significantly improved from semantic context and CS separately and in combination. In contrast, nonnative listeners could integrate sentence context information only when hearing CS sentences. The nonnative listeners were thus able to rely on top-down knowledge, just as native listeners did, but they needed enhanced signal clarity to take advantage of the contextual information. The smaller CS and NAS benefit found in the presence of SSN here could thus be rooted in the dual challenge faced by nonnative listeners: lack of experience in using the L2-specific acoustic–phonetic CS and NAS enhancements coupled with the less efficient use of signal-independent information to compensate for the signal degradation. However, the results from the 6T babble masker condition, in which nonnative listeners showed larger CS and NAS intelligibility improvement at the same SNR, suggest that our understanding of the speaking style intelligibility benefit in noise for the nonnative listeners may be incomplete. It is not likely that the source of this result lies exclusively in the largest processing difficulty in 6T babble leading to the biggest CS and NAS improvements, as we would then expect the two listener groups to show similar results. It remains a pressing goal to specify further how CS and NAS aid integration of bottom-up and top-down information and speech processing in energetic and informational maskers for native and nonnative listeners.

Conclusions

The current results indicate that the well-documented intelligibility benefit derived from the speaking style modifications crucially interacts with the masking conditions and with the listener characteristics. While many are known, the current work highlights that many aspects of these interactions remain poorly understood. The findings have implications for the testing of human speech processing in noise in clinical settings, which are typically limited to providing information about listeners' ability to understand speech in only one of many listening contexts. This study adds to the research suggesting that listeners should be tested in several maskers for a more comprehensive picture of one's ability to understand speech in noise (Van Engen, 2012). With better understanding of how CS and NAS facilitate speech processing in different listening conditions and for different listener groups, the more specific information on how to use this speech style effectively can be incorporated into training, clinical, and educational settings (see Peng & Wang, 2016). Finally, the findings reported here provide support for speech-oriented, behavioral therapy techniques, such as using CS elicited in combination with background noise (NAS+CS) rather than targeting individual speech parameters (e.g., speaking slowly) as valid approaches for maximizing speech intelligibility (Beukelman et al., 2002; Park et al., 2016; Stipancic et al., 2016; Tjaden, Sussman, & Wilding, 2014).

Acknowledgments

The authors would like to thank The University of Texas at Austin Sound Lab research assistants for their assistance with data collection.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6*(2), 81–96. https://doi.org/10.1017/s1366728903001056

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beukelman, D. R., Fager, S., Ullman, C., Hanson, E. K., & Logemann, J. (2002). The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. *Journal of Medical Speech-Language Pathology, 10*(4), 237–242.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America, 112*(1), 272–284. https://doi.org/10.1121/1.1487837

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America, 131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109. https://doi.org/10.1007/0-387-22794-6_17

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America, 120*(6), 4007–4018. https://doi.org/10.1121/1.2363929

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527–2538. https://doi.org/10.1121/1.1408946

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America, 128*(2), 860–869. https://doi.org/10.1121/1.3458857

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42. https://doi.org/10.1017/S0142716406060024

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Cooke, M., & Garcia Lecumberri, M. L. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America, 132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414–427. https://doi.org/10.1121/1.2804952

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language, 28*(2), 543–571. https://doi.org/10.1016/j.csl.2013.08.003

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America, 128*(4), 2059–2069. https://doi.org/10.1121/1.3478775

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication, 55*(4), 572–585. https://doi.org/10.1016/j.specom.2013.01.001

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America, 124*(2), 1264–1268. https://doi.org/10.1121/1.2946707

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34,* 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Durlach, N. I. (2006). Auditory masking: Need for improved conceptual structure. *The Journal of the Acoustical Society of America, 120*(4), 1787–1790. https://doi.org/10.1121/1.2335426

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target–masker similarity. *The Journal of the Acoustical Society of America, 114*(1), 368–379. https://doi.org/10.1121/1.1577562

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *International Journal of Aviation Psychology, 14*(3), 313–334. https://doi.org/10.1207/s15327108ijap1403_6

Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America, 111*(5), 2242–2249. https://doi.org/10.1121/1.1466873

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America, 116*(4), 2365–2373. https://doi.org/10.1121/1.1788730

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 55*(3), 779–790. https://doi.org/10.1044/1092-4388(2011/10-0342)

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America, 88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America, 115*(5), 2246–2256. https://doi.org/10.1121/1.1689343

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *The Journal of the Acoustical Society of America, 121*(2), 1040–1046. https://doi.org/10.1121/1.2427117

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445–2454. https://doi.org/10.1121/1.2180210

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Gilbert, R. C., Chandrasekaran, B., & Smiljanić, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America, 135*(1), 389–399. https://doi.org/10.1121/1.4838975

Glave, R. D., & Rietveld, A. C. M. (1975). Is the effort dependence of speech loudness explicable on the basis of acoustical cues? *The Journal of the Acoustical Society of America, 58*(4), 875–879. https://doi.org/10.1121/1.380737

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2013). Approaching speech intelligibility enhancement with inspiration from Lombard

and Clear speaking styles. *Computer Speech & Language, 28*(2), 629–647. https://doi.org/10.1016/j.csl.2013.09.007
Goy, H., Pichora-Fuller, M. K., & van Lieshout, P. (2013). Effects of intra-talker differences on speech understanding in noise by younger and older adults. *Canadian Acoustics, 41*(2), 23–32.
Hazan, V., Grynpas, J., & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions? *The Journal of the Acoustical Society of America, 132*(5), EL371–EL377. https://doi.org/10.1121/1.4757698
Hazan, V., & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects. *Language and Speech, 43*(3), 273–294. https://doi.org/10.1177/00238309000430030301
Heumann, C., Schomaker, M., & Shalabh. (2016). *Introduction to statistics and data analysis*. Springer. https://doi.org/10.1007/978-3-319-46162-5
Kilman, L., Zekveld, A. A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology, 5,* 651. https://doi.org/10.3389/fpsyg.2014.00651
Krause, J. C., & Braida, L. D. (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. *The Journal of the Acoustical Society of America, 125*(5), 3346–3357. https://doi.org/10.1121/1.3097491
Lenth, R. V. (2018). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.3.1. https://cran.r-project.org/package=emmeans
Liu, S., & Zeng, F.-G. (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America, 120*(1), 424–432. https://doi.org/10.1121/1.2208427
Lombard, E. (1911). The sign of the elevation of the voice. *Annales Des Maladies de l'Oreille, Du Larynx, Du Nez et Du Pharynx, 37*(2), 101–119.
Lou, H., & Ou, D. (2020). The effects of speech intelligibility on English scientific literature reading in Chinese open-plan offices. *The Journal of the Acoustical Society of America, 147*(1), EL1–EL6. https://doi.org/10.1121/10.0000497
Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*(5), 3261–3275. https://doi.org/10.1121/1.2990705
Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication, 51*(12), 1253–1262. https://doi.org/10.1016/j.specom.2009.07.002
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006
Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3), 686–693. https://doi.org/10.1044/jslhr.4003.686
Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America, 75*(2), 632–634. https://doi.org/10.1121/1.390495
Park, S., Theodoros, D., Finch, E., & Cardell, E. (2016). Be Clear: A New Intensive Speech Treatment for Adults With Nonprogressive Dysarthria. *American Journal of Speech-Language Pathology, 25*(1), 97–110. https://doi.org/10.1044/2015_AJSLP-14-0113
Peng, E., & Wang, L. M. (2016). Effects of noise, reverberation and foreign accent on native and non-native listeners' performance of English speech comprehension. *The Journal of the Acoustical Society of America, 139*(5), 2772–2783. https://doi.org/10.1121/1.4948564
Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking Clearly for the Hard of Hearing I. *Journal of Speech and Hearing Research, 28*(1), 96–103. https://doi.org/10.1044/jshr.2801.96
Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech. *Journal of Speech and Hearing Research, 29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434
Pichora-Fuller, M. K., Goy, H., & Van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. *Seminars in Hearing, 31*(2), 077–094. https://doi.org/10.1055/s-0030-1252100
Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America, 57,* S5. https://doi.org/10.1121/1.1995329
Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27,* 465–485. https://doi.org/10.1017/S014271640606036X
Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America, 133*(4), 2431–2443. https://doi.org/10.1121/1.4794379
Rönnberg, J., Lunner, T., Zekveld, A. A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7,* 1–17. https://doi.org/10.3389/fnsys.2013.00031
Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America, 118*(5), 2775–2778. https://doi.org/10.1121/1.2062650
Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100*(4), 2471–2485. https://doi.org/10.1121/1.417955
Smiljanić, R. (2013). Can older adults enhance the intelligibility of their speech? *The Journal of the Acoustical Society of America, 133*(2), EL129–EL135. https://doi.org/https://doi.org/10.1121/1.4776774
Smiljanić, R. (2021). Clear speech perception: Linguistic and cognitive benefits. In D. B. Pisoni, R. E. Remez, L. C. Nygaard, & J. S. Pardo (Eds.), *The handbook of speech perception* (2nd ed., pp. 177–205). Wiley. https://doi.org/10.1002/9781119184096.ch7
Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America, 130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Smiljanić, R., & Gilbert, R. C. (2017a). Acoustics of clear and noise-adapted speech in children, young, and older adults. *Journal of Speech, Language, and Hearing Research, 60*(11), 3081–3096. https://doi.org/10.1044/2017_JSLHR-S-16-0130

Smiljanić, R., & Gilbert, R. C. (2017b). Intelligibility of noise-adapted and clear speech in child, young adult, and older adult talkers. *Journal of Speech, Language, and Hearing Research, 60*(11), 3069–3080. https://doi.org/10.1044/2017_JSLHR-S-16-0165

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11*(2), 333–341. https://doi.org/10.1037/0882-7974.11.2.333

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America, 101*(4), 2278–2288. https://doi.org/10.1121/1.418208

Stipancic, K. L., Tjaden, K., & Wilding, G. E. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research, 59*(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271

Takata, T., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America, 88*(2), 663–666. https://doi.org/10.1121/1.399769

Tjaden, K., Kain, A., & Lam, J. (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's Disease. *Journal of Speech, Language, and Hearing Research, 57*(4), 1191–1205. https://doi.org/10.1044/2014_JSLHR-S-13-0086

Tjaden, K., Lam, J., & Wilding, G. (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research, 56*(5), 1485–1502. https://doi.org/10.1044/1092-4388(2013/12-0259)

Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication, 52*(11–12), 943–953. https://doi.org/10.1016/j.specom.2010.05.002

Van Engen, K. J. (2012). Speech-in-speech recognition: A training study. *Language and Cognitive Processes, 27*(7–8), 1089–1107. https://doi.org/10.1080/01690965.2012.654644

Van Engen, K. J. (2017). Clear speech and lexical competition in younger and older adult listeners. *The Journal of the Acoustical Society of America, 142*(2), 1067–1077. https://doi.org/10.1121/1.4998708

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121*(1), 519–526. https://doi.org/10.1121/1.2400666

Van Engen, K. J., Phelps, J. E. B., Smiljanić, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research, 57*(5), 1908–1918. https://doi.org/10.1044/JSLHR-H-13-0076

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*(3), 917–928. https://doi.org/10.1121/1.396660

Van Wijngaarden, S., Steeneken, H., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America, 111*(4), 1906–1916. https://doi.org/10.1121/1.1456928

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Zeng, F. G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research, 49*(2), 367–380. https://doi.org/10.1044/1092-4388(2006/029)

Zinszer, B. D., Riggs, M., Reetzke, R., & Chandrasekaran, B. (2019). Error patterns of native and non-native listeners' perception of speech in noise. *The Journal of the Acoustical Society of America, 145*(2), EL129–EL135. https://doi.org/10.1121/1.5087271