Perceptual consequences of native and non-native clear speech

Misaki Kato and Melissa M. Baese-Berk

Citation: The Journal of the Acoustical Society of America **151**, 1246 (2022); doi: 10.1121/10.0009403
View online: https://doi.org/10.1121/10.0009403
View Table of Contents: https://asa.scitation.org/toc/jas/151/2
Published by the Acoustical Society of America

JASA ARTICLE

Perceptual consequences of native and non-native clear speech

Misaki Kato and Melissa M. Baese-Berk[a)]
Department of Linguistics, University of Oregon, Eugene, Oregon 97403, USA

ABSTRACT:
Native talkers are able to enhance acoustic characteristics of their speech in a speaking style known as "clear speech," which is better understood by listeners than "plain speech." However, despite substantial research in the area of clear speech, it is less clear whether non-native talkers of various proficiency levels are able to adopt a clear speaking style and if so, whether this style has perceptual benefits for native listeners. In the present study, native English listeners evaluated plain and clear speech produced by three groups: native English talkers, non-native talkers with lower proficiency, and non-native talkers with higher proficiency. Listeners completed a transcription task (i.e., an objective measure of the speech intelligibility). We investigated intelligibility as a function of language background and proficiency and also investigated the acoustic modifications that are associated with these perceptual benefits. The results of the study suggest that both native and non-native talkers modulate their speech when asked to adopt a clear speaking style, but that the size of the acoustic modifications, as well as consequences of this speaking style for perception differ as a function of language background and language proficiency.
 https://doi.org/10.1121/10.0009403

(Received 7 April 2021; revised 23 November 2021; accepted 15 January 2022; published online 23 February 2022)
[Editor: Benjamin V Tucker] Pages: 1246–1258

I. INTRODUCTION

One of the most common communication difficulties that talkers face when communicating in their non-native language is that their listeners do not understand their speech. Previous research suggests that native talkers' effort to enhance acoustic-phonetic properties of their speech (i.e., clear speech enhancements) results in robust increases in understanding for listeners of various characteristics, including listeners with hearing impairments and non-native listeners (e.g., Bradlow and Alexander, 2007; Picheny *et al.*, 1985, 1986; Schum, 1996; Smiljanić and Bradlow, 2009; see Smiljanić, 2021 for a review). However, much less is known about how non-native talkers' clear speech enhancements are perceived by native listeners. Particularly, it is not clear whether non-native talkers' ability to increase intelligibility of their speech changes as their second language (L2) proficiency develops.

A. Intelligibility benefits of clear speech enhancements

Perceptual consequences of clear speech enhancements have typically been measured using an intelligibility task, where listeners hear speech materials produced in plain- and clear-speaking styles with noise and transcribe or repeat what they hear (e.g., Bradlow and Bent, 2002). Previous studies have found robust intelligibility gains resulting from native English talkers' clear speech enhancements for native English listeners of various characteristics, including hearing-impaired listeners (e.g., Ferguson, 2004; Ferguson and Morgan, 2018; Krause and Braida, 2002; Liu *et al.*, 2004; Picheny *et al.*, 1985; Schum, 1996; Smiljanić and Bradlow, 2005; Uchanski *et al.*, 1996). Furthermore, native English talkers' clear speech enhancements result in intelligibility gains for non-native English listeners (e.g., Bradlow and Bent, 2002; Bradlow and Alexander, 2007). A similar intelligibility benefit has also been reported for clear speech in languages other than English, including Croatian and French (Gagné *et al.*, 1994; Gagné *et al.*, 2002; Smiljanić and Bradlow, 2005).

While clear speech enhancements made in talkers' native languages have been shown to result in reliable intelligibility gains for a variety of listeners, much less is known about how clear speech enhancements made in a non-native language are perceived by native listeners. Specifically, existing literature regarding intelligibility gains resulting from non-native talkers' clear speech enhancements is mostly limited to those produced by highly proficient non-native talkers. For example, speech enhancements made by highly proficient non-native talkers demonstrate similar intelligibility gains compared to clear speech produced by native talkers (Rogers *et al.*, 2010; Smiljanić and Bradlow, 2005, 2011). Further, the size of acoustic modifications made by native talkers and by highly proficient non-native talkers are similar (Granlund *et al.*, 2012). However, data from talkers of lower proficiency are relatively scarce. One study demonstrated that late learners of English were much less effective at enhancing intelligibility of English vowels than early learners and native English talkers (Rogers *et al.*, 2010). Specifically, clear speech enhancements of English vowels in /bVd/ syllables produced by monolingual native English talkers and early learners of English resulted in

[a)]Electronic mail: mbaesebe@uoregon.edu

 0001-4966/2022/151(2)/1246/13/$30.00

similar intelligibility gains; however, those produced by late learners resulted in much smaller intelligibility gains. In fact, the late learners' clear speech enhancements sometimes resulted in a decrease in intelligibility.

These studies suggest that non-native talkers' target language proficiency may affect their ability to increase intelligibility. That is, the more familiar the talkers are with the sound structure of the language, including the system of phonological contrasts and phonetic implementation of those contrasts, the more effective their clear speech enhancements may be at increasing intelligibility for native listeners (Smiljanić and Bradlow, 2011). However, with the data from existing literature, it is difficult to determine the effect of target language proficiency on clear speech enhancements beyond the level of single sound production (see, e.g., Kato and Baese-Berk, 2021). That is, while the more experienced L2 learners are better able to increase the intelligibility of English vowels than less experienced L2 learners (Rogers *et al.*, 2010), it is not clear whether the effect of target language experience generalizes to clear speech intelligibility benefits for sentence production. Increasing intelligibility of sentences can be more challenging than doing so for single words because it requires proficient use of the target language sound system at multiple levels, including single sounds or words, in addition to other features such as prosody. Thus, the effect of non-native talkers' proficiency level on clear speech intelligibility benefits could manifest differently at the sentence level compared to the single-word level.

Furthermore, examining the intelligibility of the speech produced in plain- and clear-speaking styles by non-native talkers of different proficiency levels may help us better understand the relationship between talkers' ability to produce intelligible speech in general vs their ability to *increase* intelligibility of their speech. Specifically, given that producing speech in a non-native language becomes more fluent and less effortful as the talkers' proficiency develops (e.g., Kormos and Dénes, 2004; Nip and Blumenfeld, 2015), it is possible that higher-proficiency non-native talkers' speech (their speech in plain- and clear-speaking styles) is generally more intelligible than lower-proficiency talkers' speech. However, the ability to further *increase* intelligibility (i.e., intelligibility improvement from plain to clear speech) may or may not differ between non-native talkers of differing proficiency levels. That is, if higher-proficiency talkers are able to increase intelligibility of their speech to a larger extent than lower-proficiency talkers, this would suggest that non-native talkers' increased proficiency is associated with their ability to not only produce generally more intelligible speech but also with their ability to further *increase* intelligibility of their speech. However, if the size of intelligibility improvement is similar between the speech of higher- and lower-proficiency talkers, it may suggest that the ability to produce generally intelligible speech and the ability to increase intelligibility are at least partially independent from one another. In order to answer these questions, we examine clear speech intelligibility benefits of English sentences produced by native English talkers and non-native talkers of higher- and lower-proficiency.

B. Current study

In the current study, we examine intelligibility benefits resulting from clear speech enhancements produced by native and non-native talkers of English. Specifically, we ask whether native and non-native talkers' clear speech enhancements result in a similar size of intelligibility improvement for native English listeners. Further, we ask whether clear speech produced by higher-proficiency non-native talkers results in a larger intelligibility improvement compared to that produced by lower-proficiency non-native talkers. Given that non-native talkers' phonological representations of non-native sounds may not be the same as those of native talkers (Imai *et al.*, 2005), non-native talkers may emphasize different acoustic cues than native talkers to enhance intelligibility of non-native sounds. Thus, clear speech modifications made by non-native talkers (including higher- and lower-proficiency talkers) may be much smaller than those made by native talkers for native listeners. However, given previous results demonstrating that clear speech enhancements of English vowels produced by early L2 learners resulted in larger intelligibility gains than those produced by late L2 learners (Rogers *et al.*, 2010), it is also possible that higher-proficiency talkers' intelligibility improvement of sentences is much larger than that of lower-proficiency talkers.

Below, we describe the acoustic-phonetic modifications made by three groups of talkers (native English talkers, and higher- and lower-proficiency native Mandarin talkers) when producing clear and plain speech. Following this discussion, we describe the results of an experiment designed to investigate intelligibility of plain and clear speech produced by these three groups of talkers.

II. DESCRIPTION OF ACOUSTIC CHARACTERISTICS OF THE STIMULI USED IN THE INTELLIGIBILITY PERCEPTION EXPERIMENT

A. Methods

1. Materials

The test materials were 30 sentences chosen from the Revised Bamford-Kowal-Bench Standard Sentence Test (BKB sentences; Bamford and Wilson, 1979), developed by the Cochlear Corporation for use with American children. They were simple English sentences, each sentence consisting of three or four keywords (e.g., *The shop closed for lunch*). These sentences are provided in the Appendix.

2. Participants (talkers)

Four native English talkers (age range = 19–22 years, mean = 20) and 8 non-native English talkers whose native language was Mandarin Chinese (age range = 20–31 years, mean = 25.3) recorded the sentences. All talkers identified themselves as female and reported no history of speech or hearing impairment. Native English talkers were recruited from the University of Oregon Psychology and Linguistics human subject pool, and they received partial course credit for their participation. We recruited non-native English talkers from two different instructional settings. Specifically, we recruited four higher-proficiency non-native talkers (Native Mandarin-High talkers) from the graduate student population at the University of Oregon, and four lower-proficiency non-native talkers (Native Mandarin-Low talkers) from an intensive English program, who were international students studying English before entering the university as matriculated students. In order to ensure that higher- and lower-proficiency non-native talkers examined here are of different proficiency levels, we collected information about non-native talkers' English proficiency from multiple sources, including the information collected from a language background questionnaire (e.g., information about length of residence in the English-speaking country, standardized English proficiency test score; see Table I) as well as non-native talkers' perceived accentedness (evaluated by native English listeners; see supplementary material).[1]

3. Procedure

All talkers were recorded in a sound booth. The sentences were displayed on the computer screen one at a time; the presentation of each sentence was self-paced. Recording was done on a single channel at a sampling rate of 44 100 Hz (16 bit) using the Praat speech analysis software package (Boersma and Weenink, 2018). Talkers were instructed to read the test sentences in a plain-speaking style first, followed by the second recording in a clear-speaking style. For the recordings in the plain-speaking style, the talkers were instructed to read as if they were talking to someone who is familiar with their voice and speech patterns. For the recordings in the clear-speaking style, the talkers were instructed to read as if they were talking to a listener who has a hearing loss (Smiljanić and Bradlow, 2011). After the recording, talkers completed a language background questionnaire. All speech files were segmented into individual sentence-length files.

4. Measurements

We examined characteristics of the materials in acoustic analyses. For each test sentence in each of the plain- and clear-speaking styles for each talker, we examined speaking rate, mean $F0$, and $F0$ range, and vowel space, which are acoustic features typically examined in previous studies (e.g., Bradlow *et al.*, 2003; Granlund *et al.*, 2012; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2005). In order to examine how temporal characteristics differ between plain and clear speech produced by native talkers and non-native talkers of different proficiency levels, speaking rate was computed by dividing the number of syllables of the sentence by the sentence duration (in seconds). Raw speaking rate (including within-sentence pauses)[2] was examined to explore whether there are differences based on the talker group (e.g., English proficiency of the talkers) in overall speaking rate. Further, in order to examine whether talkers made a difference in speaking rate between plain- and clear-speaking styles, scaled values of speaking rate were also computed using the min-max scaling procedure (Gerstman, 1968; Kallay and Redford, 2018). That is, speaking rate for a particular sentence was normalized using the talker's minimum and maximum values of speaking rate so that all the values are within the range of 0 (minimum value of that talker) to 1 (maximum value of that talker).

In order to examine how plain-clear style differences are manifested in fundamental frequency, we measured mean $F0$ and $F0$ range. A PRAAT script was run to calculate mean $F0$, maximum $F0$, and minimum $F0$ (in Hertz) for each sentence. $F0$ range was obtained by subtracting the minimum $F0$ value from the maximum $F0$ value for each sentence. The values of mean $F0$ and $F0$ range were transformed using the min-max scaling procedure in order to examine the plain-clear style difference and whether it differed across talker groups.

Finally, we examined how plain-clear style differences are manifested in the vowel space of the native and non-native English speakers' productions. That is, we examined whether the vowel space area would differ in plain vs clear speech as one way to capture overall differences in vowel characteristics (e.g., Krause and Braida, 2004; Smiljanić and Bradlow, 2005). We selected 4 point-vowels in order to

TABLE I. Non-native English (native Mandarin) talkers' English learning background and proficiency.

Talker	Age	Age of onset for English speaking	Years of formal English training	Length of US residence in months	TOEFL score
Native Mandarin-low (NM-L): average	20.5	15.5	7	18.8	45.3
NM-L 103	21	19	9	24	52
NM-L 104	20	15	6	15	35
NM-L 106	20	13	7	19	53
NM-L 107	21	15	6	17	41
Native Mandarin-high (NM-H): average	30	17.5	15.3	62.5	93.8
NM-H 302	31	23	12	27	108
NM-H 306	30	13	9	108	91
NM-H 310	28	10	15	19	106
NM-H 311	31	24	25	96	70

characterize each talker's vowel space for clear and plain speech: /i/, /æ/, /ɑ/ and /u/ (Ferguson and Kewley-Port, 2007). Phone-level alignment between sound files and transcripts of the sentence was automated using Montreal Forced Aligner (McAuliffe *et al.*, 2017). Then, automated vowel formant extraction was carried out using Forced Alignment and Vowel Extraction (Rosenfelder *et al.*, 2014). *F1* and *F2* frequencies of the 4 point-vowels were taken from the midpoint (i.e., 50% of the vowel duration) of each vowel. Midpoint *F1* and *F2* were then z-score normalized to control for individual differences (i.e., Lobanov method: Nearey, 1977; Thomas and Kendall, 2015). Vowel space area was measured as the Euclidean area covered by the quadrilateral defined by the mean of each of 4 point-vowels, using R package phonR (McCloy, 2016). Vowel space was calculated for each speaking style (plain and clear) for each talker.

B. Results

1. Speaking rate

The left panel in Fig. 1 shows raw speaking rate (i.e., number of syllables divided by the sentence duration in seconds) for Native English, Native Mandarin-High and Native Mandarin-Low talkers' productions. A one-way analysis of variance (ANOVA)[3] examining the effect of talker groups (i.e., Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) on the raw speaking rate was conducted using the afex package (Singmann *et al.*, 2020) within the R computing program. The results showed a significant effect of talker group [$F(2, 9) = 24.34$, $p < 0.001$, $\eta^2_p = 0.84$, $\eta^2_G = 0.84$].[4] A *post hoc* Tukey test revealed that speaking rate was faster for Native English talkers than for Native Mandarin-High talkers ($\beta = 1.2$, standard error, $SE = 0.23$, t ratio $= 5.25$, $p = 0.001$), but did not significantly differ for the speech of Native Mandarin-High talkers and Native Mandarin-Low talkers ($\beta = 0.31$, $SE = 0.23$, t ratio $= 1.35$, $p = 0.4$). This confirmed that Native English talkers generally spoke faster than Native Mandarin talkers (both High and Low), but Native Mandarin-High talkers did not speak significantly faster than Native Mandarin-Low talkers.

In the next analysis, we examined whether talkers produced speech more slowly in the clear-speaking style than in the plain-speaking style, and whether this pattern differed for different talker groups' speech. The right panel in Fig. 1 (the scaled values of the speaking rate) suggests that Native English and Native Mandarin-High talkers spoke more slowly in the clear-speaking style than in the plain-speaking style, but this difference is much smaller for Native Mandarin-Low talkers' speech. A two-way within-subjects ANOVA was conducted examining the effect of speaking style (i.e., clear or plain) on the scaled speaking rate, and whether the effect of speaking style differed for the speech of different talker groups (i.e., Native English, Native Mandarin-High, and Native-Mandarin-Low talkers). Because each talker produced both speaking styles, speaking style was treated as a within-subject factor. The results showed a significant effect of speaking style [$F(1, 9) = 111.07$, $p < 0.001$, $\eta^2_p = 0.93$, $\eta^2_G = 0.82$], as well as a significant interaction between speaking style and talker group [$F(2, 9) = 11.12$, $p = 0.004$, $\eta^2_p = 0.71$, $\eta^2_G = 0.47$]. *Post hoc* Tukey pairwise comparisons between speaking styles within each talker group confirmed that the plain-clear style difference in the scaled articulation rate was significant for the Native English ($\beta = 0.36$, $SE = 0.05$, t ratio $= 7.99$, $p < 0.0001$) and Native Mandarin-High group ($\beta = 0.36$, $SE = 0.05$, t ratio $= 8.02$, $p < 0.0001$), but not for Native Mandarin-Low group ($\beta = 0.1$, $SE = 0.04$, t ratio $= 2.3$, $p = 0.052$). In order to further examine the interaction between speaking style and talker group, a two-way within-subject ANOVA was conducted for subsets of the data: for Native English and Native Mandarin-High talkers' data, and for Native Mandarin-High

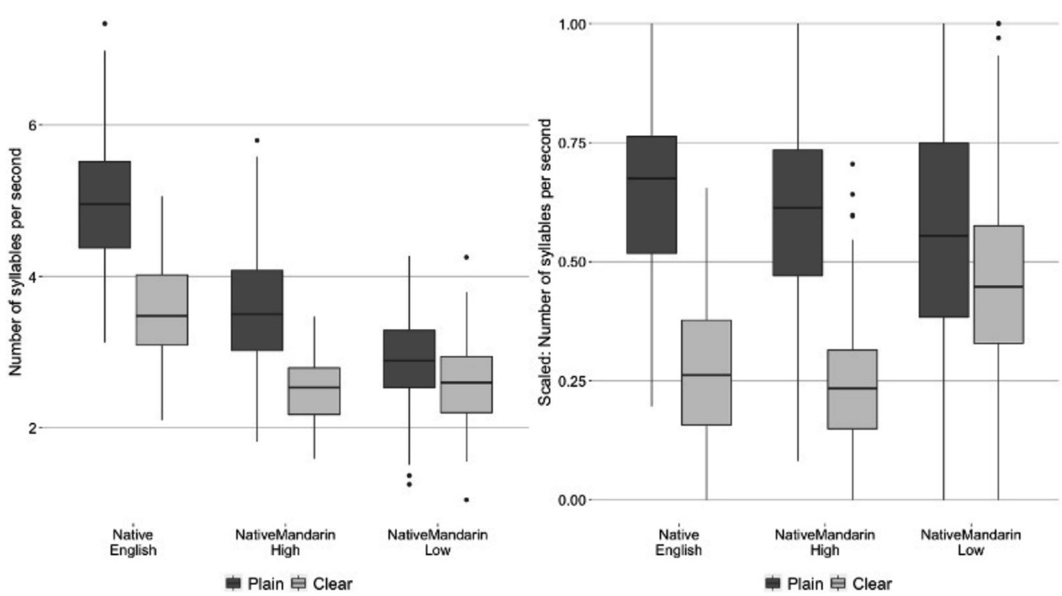

FIG. 1. Left panel: Raw speaking rate for the speech produced by different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in two speaking styles (Plain and Clear). Right panel: Scaled speaking rate for different talker groups' speech in two speaking styles.

and Native Mandarin-Low talkers' data. These tests revealed that the speaking style x talker group interaction was significant for the Native Mandarin-High vs Native Mandarin-Low talker group comparison [$F(1, 6) = 14.96$, $p = 0.008$, $\eta^2_{\mathrm{p}} = 0.71$, $\eta^2_{\mathrm{G}} = 0.46$], but not for the Native English vs Native Mandarin-High talker group comparison [$F(1, 6) = 0.00$, $p = 0.99$, $\eta^2_{\mathrm{p}} < 0.001$, $\eta^2_{\mathrm{G}} < 0.001$]. Together, these results demonstrated different patterns for overall speaking rate and for plain-clear style differences in speaking rate. That is, in terms of the overall speaking rate, Native English talkers spoke faster than Native Mandarin talkers, but Native Mandarin-High talkers did not speak faster than Native Mandarin-Low talkers. In terms of the plain-clear style differences in speaking rate, Native English and Native Mandarin-High talkers made a larger difference between plain- and clear-speaking styles than Native Mandarin-Low talkers did. The Native English and Native Mandarin-High talkers slowed down their speaking rate from the plain to clear speaking style to a similar extent (the mean number of syllables per second in plain speech–clear speech for Native English, 1.43; for Native Mandarin-High, 1.1). This suggests that overall speaking rate is partially independent from clear speech modifications in speaking rate.

2. Fundamental frequency

Figure 2 shows the scaled values of $F0$ range and mean $F0$ for the sentences produced by Native English, Native Mandarin-High, and Native-Mandarin-Low talkers in plain- and clear-speaking styles. The figure suggests that there was a general trend for clear-style sentences to have wider $F0$ range and higher mean $F0$ than for plain-style sentences across the three talker groups. However, the difference between the two speaking styles is smaller for the sentences produced by the Native-Mandarin-Low talkers compared to those produced by Native English and Native Mandarin-High talkers.

In order to examine the effect of the speaking style (i.e., plain or clear) and whether it differs for different talker groups, two sets of two-way within-subjects ANOVAs were conducted with the scaled values of $F0$ range and mean $F0$ as the dependent variable. ANOVA results for $F0$ range showed a significant effect of speaking style [$F(1, 9) = 16.86$, $p = 0.003$, $\eta^2_{\mathrm{p}} = 0.65$, $\eta^2_{\mathrm{G}} = 0.26$] as well as a significant interaction between speaking style and talker group [$F(2, 9) = 5.06$, $p = 0.034$, $\eta^2_{\mathrm{p}} = 0.53$, $\eta^2_{\mathrm{G}} = 0.17$], but not a significant effect of talker group [$F(2, 9) = 2.39$, $p = 0.15$, $\eta^2_{\mathrm{p}} = 0.35$, $\eta^2_{\mathrm{G}} = 0.3$]. *Post hoc* Tukey pairwise comparisons examined the effect of speaking style within each talker group; the tests revealed a significant plain-clear style difference for the Native English ($\beta = -0.16$, SE $= 0.05$, t ratio $= -3.58$, $p = 0.0059$) and Native Mandarin-High groups ($\beta = -0.17$, SE $= 0.05$, t ratio $= -3.75$, $p = 0.0045$), but not for the Native Mandarin-Low group ($\beta = 0.01$, SE $= 0.05$, t ratio $= 0.23$, $p = 0.83$). In order to further examine the interaction between speaking style and talker group, a two-way within-subject ANOVA was conducted for subsets of the data. These tests revealed that the speaking style x talker group interaction was significant for the Native Mandarin-High vs Native Mandarin-Low talker group comparison [$F(1, 6) = 6.04$, $p = 0.049$, $\eta^2_{\mathrm{p}} = 0.5$, $\eta^2_{\mathrm{G}} = 0.26$], but not for the Native English vs Native Mandarin-High talker group comparison [$F(1, 6) = 0.02$, $p = 0.898$, $\eta^2_{\mathrm{p}} = 0.003$, $\eta^2_{\mathrm{G}} < 0.001$]. Thus, the size of plain-to-clear style difference in $F0$ range was similar between Native English and Native Mandarin-High talkers' speech but was larger for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech.

For mean $F0$, the effect of speaking style was significant [$F(1, 9) = 23.63$, $p < 0.001$, $\eta^2_{\mathrm{p}} = 0.72$, $\eta^2_{\mathrm{G}} = 0.5$], but the effect of talker group was not [$F(2, 9) = 1.71$, $p = 0.24$, $\eta^2_{\mathrm{p}} = 0.28$, $\eta^2_{\mathrm{G}} = 0.19$]. The interaction between speaking style and talker group was not significant [$F(2, 9) = 1.57$,

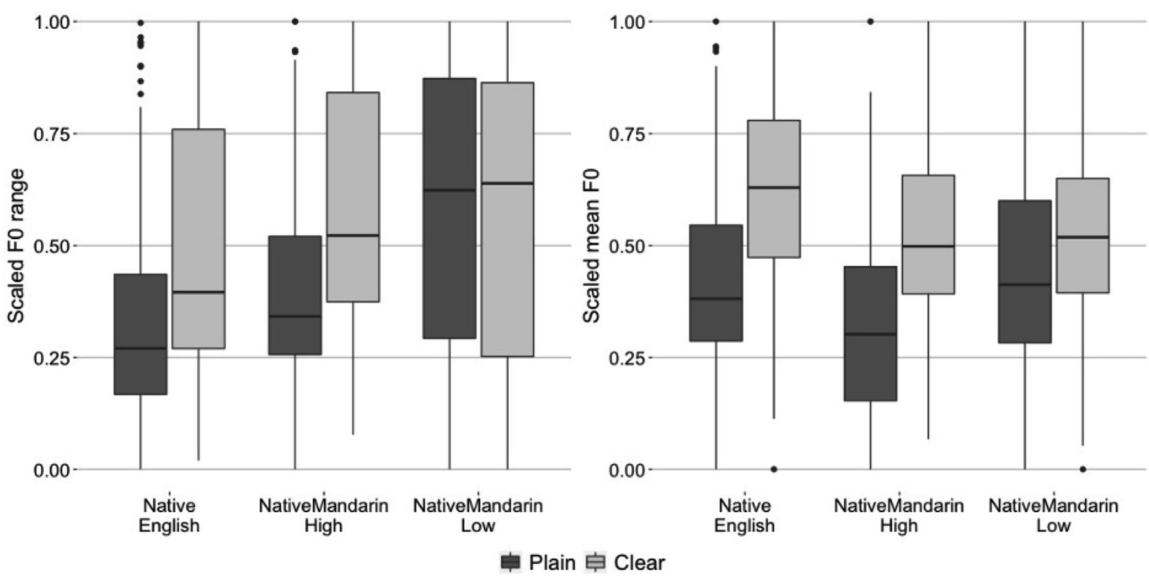

FIG. 2. Scaled $F0$ range (left panel) and scaled mean $F0$ (right panel) for Native English, Native Mandarin-High, and Native Mandarin-Low talkers' productions in two speaking styles (Plain and Clear).

https://doi.org/10.1121/10.0009403

$p = 0.26$, $\eta^2_p = 0.26$, $\eta^2_G = 0.12$]. However, *post hoc* Tukey pairwise comparisons showed a significant plain-clear difference for the Native English ($\beta = -0.21$, SE = 0.06, *t* ratio = –3.58, $p = 0.006$) and Native Mandarin-High groups ($\beta = -0.2$, SE = 0.06, *t* ratio = –3.47, $p = 0.007$), but not for the Native Mandarin-Low group ($\beta = -0.08$, SE = 0.06, *t* ratio = –1.36, $p = 0.21$). Two-way within-subject ANOVAs conducted for subsets of the data also showed a significant speaking style x talker group interaction for the Native Mandarin-High vs Native Mandarin-Low talker group comparison [$F(1, 6) = 6.34$, $p = 0.045$, $\eta^2_p = 0.51$, $\eta^2_G = 0.12$], but not for the Native English vs Native Mandarin-High talker group comparison [$F(1, 6) = 0.005$, $p = 0.95$, $\eta^2_p < 0.001$, $\eta^2_G < 0.001$]. These results showed that, though there was not a significant difference in the size of plain-clear modifications in mean $F0$ across different talker groups' speech, there was a tendency that Native English and Native Mandarin-High talkers made a significant plain-clear style difference but Native Mandarin-Low talkers did not.

Together, these results demonstrated that Native English and Native Mandarin-High talkers produced their clear speech with wider $F0$ range and higher mean $F0$ compared to their plain speech. The size of plain-clear differences for these measures was comparable between Native English and Native Mandarin-High talkers' speech. However, the plain-clear style differences in these features were much smaller for the Native Mandarin-Low talkers' speech.

3. Vowel space

Figure 3 illustrates the vowel space area, covered by the quadrilateral defined by the mean of the 4 point-vowels (/i/, /æ/, /ɑ/, and /u/), for each talker in plain- and clear-speaking styles. The figure suggests that the vowel space expansion from plain to clear speech is the largest for the Native English talkers' speech, followed by the Native Mandarin-High talkers' speech. The vowel space expansion is the smallest for the Native Mandarin-Low talkers' speech.

In order to test these observations, a two-way within-subject ANOVA was conducted with the vowel space area as the dependent variable. The effect of speaking style (plain vs clear), the effect of talker group (Native English, Native

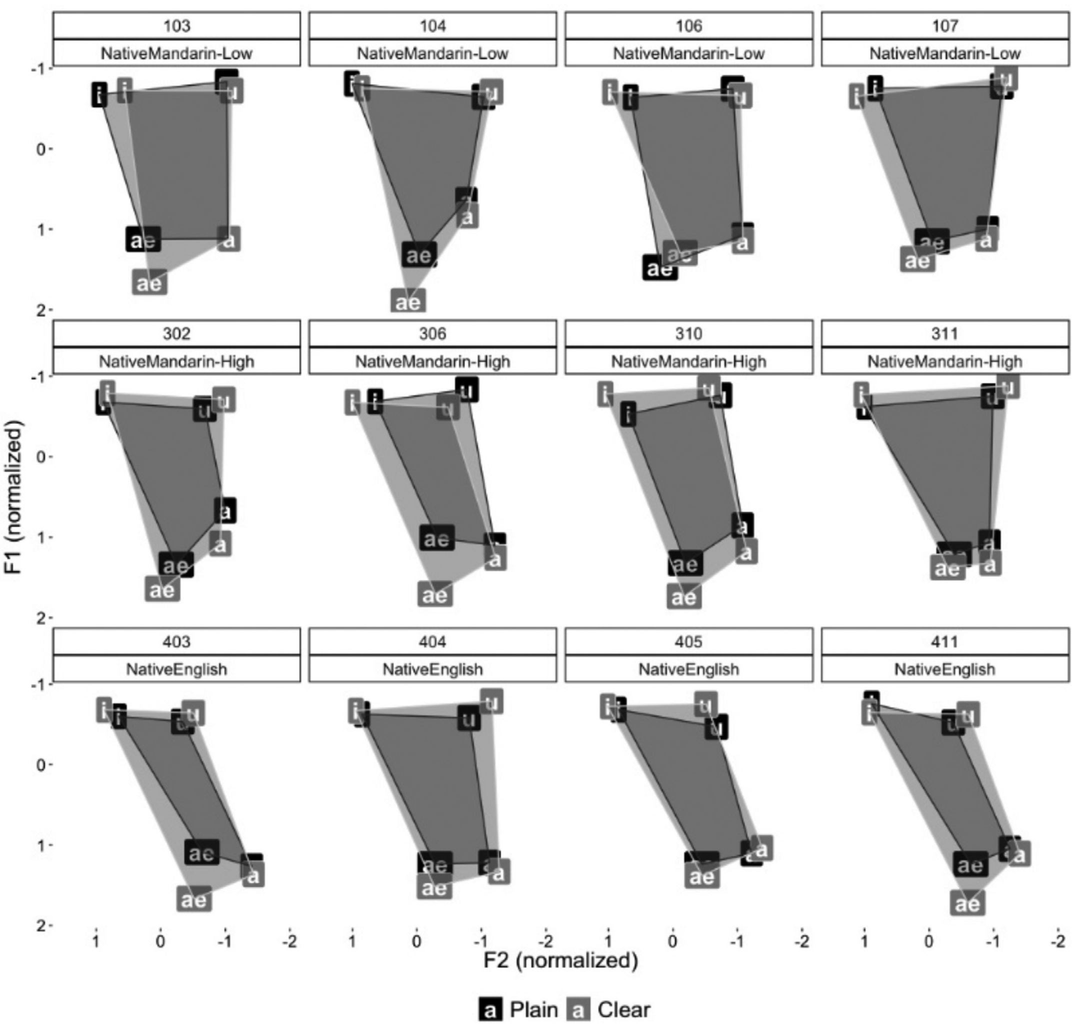

FIG. 3. Vowel space area measured as the Euclidean area covered by the quadrilateral defined by the mean of the 4 point-vowels: /i/, /æ/, /ɑ/, and /u/. Darker area is the vowel space for plain speech and lighter area is the vowel space for clear speech for each talker. The upper rows (103–107) includes Native Mandarin-Low talkers, the middle row (302–311) includes Native Mandarin-High talkers, and bottom row (403–411) includes Native English talkers.

Mandarin-High, and Native Mandarin-Low), and the interaction between the two on the vowel space area was examined. ANOVA results showed a significant effect of speaking style [$F(1, 9) = 66.68$, $p < 0.001$, $\eta^2_p = 0.88$, $\eta^2_G = 0.69$], talker group [$F(2, 9) = 6.36$, $p = 0.019$, $\eta^2_p = 0.59$, $\eta^2_G = 0.5$], and the interaction between speaking style and talker group [$F(2, 9) = 4.45$, $p = 0.045$, $\eta^2_p = 0.5$, $\eta^2_G = 0.23$]. The *post hoc* Tukey comparisons confirmed that the effect of speaking style was significant in all the talker groups' speech: Native English ($\beta = 1.05$, SE = 0.16, t ratio = 6.57, $p = 0.0001$), Native Mandarin-High ($\beta = 0.82$, SE = 0.16, t ratio = 5.16, $p = 0.0006$), Native Mandarin-Low ($\beta = 0.39$, SE = 0.16, t ratio = 2.42, $p = 0.039$). In order to further examine the interaction between speaking style and talker group (as illustrated in Fig. 4) a two-way within-subject ANOVA was conducted for subsets of the data: for Native English and Native Mandarin-High talkers' data, for Native English and Native Mandarin-Low talkers' data, and for Native Mandarin-High and Native Mandarin-Low talkers' data. These tests revealed that the speaking style x talker group interaction was significant for the Native English vs Native Mandarin-Low talker group comparison [$F(1, 6) = 6.14$, $p = 0.048$, $\eta^2_p = 0.51$, $\eta^2_G = 0.25$], but not for the Native English vs Native Mandarin-High talker group comparison [$F(1, 6) = 2.09$, $p = 0.199$, $\eta^2_p = 0.26$, $\eta^2_G = 0.04$] or for the Native Mandarin-High vs Native Mandarin-Low talker group comparison [$F(1, 6) = 3.35$, $p = 0.117$, $\eta^2_p = 0.36$, $\eta^2_G = 0.23$].

These results demonstrated that the size of vowel space expansion from plain to clear speech differed for Native English, Native Mandarin-High, and Native Mandarin-Low talkers' speech. Specifically, Native English talkers expanded their vowel space the most, followed by Native Mandarin-High and Native Mandarin-Low talkers. The vowel space area in plain and clear speech further suggested that vowel space expansion is the smallest for Native Mandarin-Low talkers' speech partly because their vowel space for plain speech was already large (i.e., Native Mandarin-Low talkers were using the vowel space that is close to their maximum in plain speech; thus there was not much room to expand in clear speech).[5]

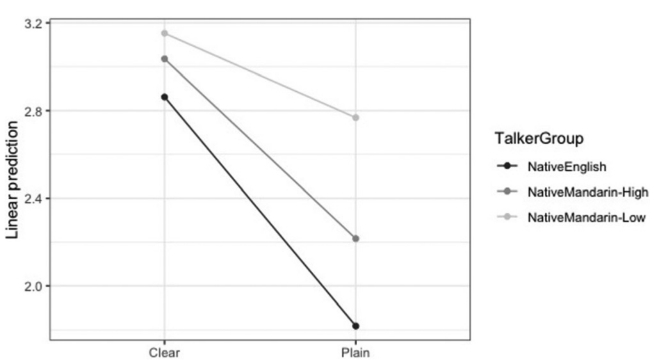

FIG. 4. Linear prediction for vowel space area for the three talker groups (Native English, Native Mandarin-High, and Native Mandarin-Low) in plain- and clear-speaking styles.

Overall, the results demonstrated that talkers generally decreased their speaking rate, increased $F0$ range, mean $F0$, and vowel space in clear speech compared to plain speech. However, there were differences in the degrees to which the talkers of different native language backgrounds/proficiency levels modified these acoustic features. That is, the native English talkers and higher-proficiency non-native talkers modified the above-mentioned acoustic features to larger degrees than lower-proficiency non-native talkers did. These results suggest that non-native talkers' clear speech strategies change as their target language proficiency develops; higher-proficiency talkers modify acoustic features to larger degrees than lower-proficiency talkers, and higher-proficiency talkers' strategies are comparable to native talkers' strategies.

III. INTELLIGIBILITY OF NATIVE AND NON-NATIVE CLEAR SPEECH

In order to understand the intelligibility benefits of native and non-native clear speech, we had native English listeners transcribe the materials (described previously) produced by native and non-native English talkers in plain- and clear-speaking styles.[6]

A. Methods

1. Participants (listeners)

Participants were 194 native English listeners (94 female, 99 male, 1 declined to provide a gender; mean age = 35.9 years). Participants were recruited using Amazon Mechanical Turk. None of the listeners reported a history of speech or hearing impairment. All participants resided in the United States, and self-reported to be native speakers of American English. None of the participants reported experience with Mandarin Chinese. Additional native English listeners were recruited for evaluating perceived accentedness of the native and non-native English talkers who recorded the materials (see supplementary material).[1]

2. Materials

Materials were 720 unique sound files (as described in Sec. II): 30 BKB sentences × 2 speaking styles (Plain and Clear) × 12 talkers (4 Native English talkers + 4 Native Mandarin-High talkers + 4 Native Mandarin-Low talkers). These files were root-mean-square (RMS) normalized to 65 dB SPL. Silence of 500 ms was then added at the beginning and end of each sound file. Further, in order to create materials for the speech-in-noise intelligibility task, we mixed each file with different sections of a speech-shaped noise file (Bradlow and Alexander, 2007) at a signal-to-noise ratio (SNR) of −6 dB for native talkers' items and −2 dB for non-native talkers' items. These SNRs were determined based on pilot testing, where we examined the noise level that would have native and non-native talkers' plain speech intelligibility to be within the range of 45%–65% correct (Smiljanić and Bradlow, 2011) so that we could assess the amount of clear speech benefit from a similar baseline level (i.e., plain speech in noise) of recognition accuracy.

3. Procedure

The experiment was conducted online using a Qualtrics link provided to participants *via* Amazon Mechanical Turk. On each trial, participants were asked to listen to an English sentence and type what they heard. They could listen to the sentence only once but could take as much time as needed to type their answer. They also completed two practice trials with talkers and sentences that were different from the 30 test sentences.

During the test trials, each participant listened to 30 unique sentences produced by six talkers (i.e., five sentences from each of two Native English, two Native Mandarin-High, two Native Mandarin-Low talkers). They heard half of the sentences (15 sentences) in a plain-speaking style and half in a clear-speaking style. Each participant heard the same number of clear- and plain-style sentences from the three talker groups. That is, they heard five clear- and five plain-style sentences produced by Native English talkers, five clear- and five plain-style sentences produced by Native Mandarin-High talkers, and five clear- and five plain-style sentences produced by Native Mandarin-Low talkers. The combination of the talker, sentence, and style was counterbalanced for each participant. The presentation order of the sentences was randomized for each participant. After the experimental trials were completed, each participant completed a post-test demographic survey.

4. Scoring and analysis

The intelligibility data were analyzed for proportion of keywords correctly recognized. Keywords were defined to be content words (e.g., *the shop closed for lunch;* see also the Appendix). Words correct were defined as those that matched the intended target exactly, as well as homophones and/or common misspellings (e.g., *to* for *too* in the sentence *The car is going too fast*). However, words with incorrect, added, or deleted morphemes were scored as incorrect (e.g., *ties* for *tied* in the sentence *The man tied his shoes*, or *shoe* for *shoes* in the same sentence). In terms of the number of the data points, there were 97 participants who listened to one set of six talkers (50 participants in the quiet condition and 47 participants in the noise condition) and 97 participants who listened to the other set of six talkers (48 participants in the quiet condition and 49 participants in the noise condition); see supplementary material[1] for the comparison of the current data set (i.e., noise condition) with the data in the quiet condition. Thus, there were 18 236 data points analyzed (194 participants × 94 keywords). The first author and a research assistant scored these data; both raters scored all of the data. When there was a disagreement (16 instances out of 18 236 instances), the two raters discussed discrepancies until they reached agreement.

B. Results

Figure 5 shows the proportion correct of keyword recognition for speech produced by different talker groups (Native English, Native Mandarin-High, and Native Mandarin-Low) in two speaking styles (plain and clear).[7] The figure suggests that listeners' keyword recognition improved from plain to clear speech for Native English (14% increase) and Native Mandarin-High talkers' speech (9% increase), but not for Native Mandarin-Low talkers' speech (2% decrease). In order to statistically examine whether listeners' keyword recognition improved for clear speech compared to plain speech and whether this effect of speaking style was different for the speech produced by different talker groups, we analyzed the intelligibility data using logistic mixed-effects regression models using R package lme4 (Bates *et al.*, 2015). The dependent variable was keyword correct, scored as a 0 for incorrect and 1 for correct for each keyword in the sentence. As fixed effects, Style (plain or clear), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and interaction between the two were included. Talker Group was also contrast-coded to compare between Native English and Native Mandarin-High group (0.5, –0.5, 0: TalkerGroup1), and between Native Mandarin-High and Native Mandarin-Low group (0, 0.5, –0.5: TalkerGroup2). Style was contrast-coded to compare between plain (–0.5) and clear (0.5) speaking styles. The maximal random effects structure that would converge was implemented, which included random intercepts for talker, listener, and item. The random effects structure also included by-talker random slope for Style, by-listener slopes for Style, Talker Group and their interaction, and by-item slopes for Style, Talker Group, and their interaction.

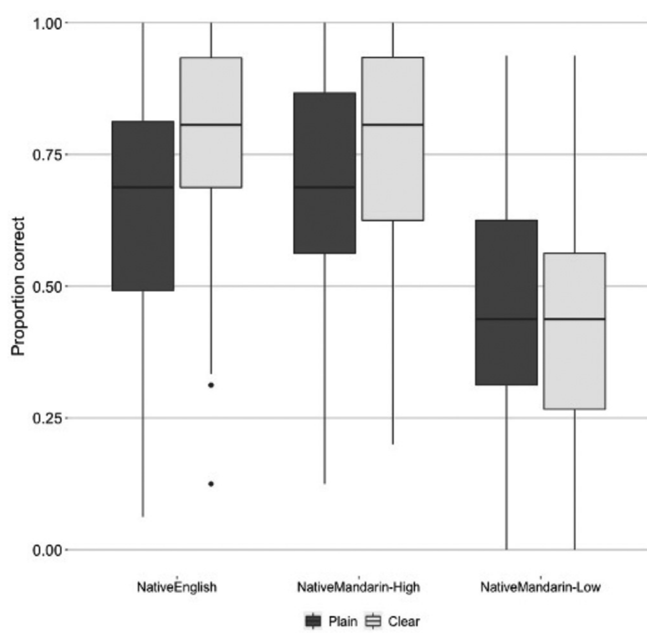

FIG. 5. Proportion correct of keyword recognition for each talker group (Native English, Native Mandarin-High, and Native Mandarin-Low) by speaking style (plain and clear).

The results showed that Native Mandarin-High talkers' speech was generally understood more accurately than Native Mandarin-Low talkers' speech (at the same noise level for these types of speech; $\beta = 2.01$, $z = 4.95$, $p < 0.001$). The results also showed a significant effect of Style ($\beta = 0.43$, $z = 2.54$, $p < 0.05$), indicating that listeners' keyword recognition accuracy was generally higher for clear speech than for plain speech. However, this effect of Style interacted with the Native Mandarin-High vs Native Mandarin-Low comparison ($\beta = 1.01$, $z = 2.34$, $p < 0.05$). This indicates that listeners' keyword recognition accuracy improved from plain to clear speech to a larger degree for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech. The effect of Style was similar between Native English and Native Mandarin-High talkers' speech (at the different noise levels used here; $\beta = 0.74$, $z = 1.71$, $p = 0.87$). To further examine the interaction between Style and Talker Group, a *post hoc* Tukey test was conducted. The test showed that the effect of Style on keyword recognition was significant for the speech produced by Native English talkers ($p = 0.004$) and for Native Mandarin-High talkers ($p = 0.045$), but not for the speech produced by Native Mandarin-Low talkers ($p = 0.79$). Thus, the intelligibility improvement from plain to clear speech was larger for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech. Indeed, there was no such intelligibility improvement for Native Mandarin-Low talkers' speech. The model summary is presented in Table II.

IV. DISCUSSION

The present study examined native English listeners' perception of clear speech produced by native English talkers and non-native English talkers of different proficiency levels. We demonstrated that listeners generally transcribed clear speech more accurately than plain speech. However, the size of intelligibility improvement from plain to clear speech differed for the speech produced by native English talkers, higher-proficiency non-native talkers, and lower-proficiency non-native talkers. Specifically, the lower-proficiency talkers did not demonstrate a plain-to-clear intelligibility improvement, but higher-proficiency talkers and native English talkers did. Interestingly, the size of intelligibility improvement did not significantly differ for the speech of higher-proficiency talkers and native English talkers.[8]

TABLE II. Summary of the mixed-effects logistic regression model for the intelligibility data in the noise condition, as well as the results of the *post hoc* Tukey test comparing the effect of Style (plain vs clear) for each talker groups' speech.

```
Response ~ Style*TalkerGroup
+ (1+ Style| Talker)
+ (1+ Style* TalkerGroup| Listener)
+ (1+ Style* TalkerGroup| Item)
```

Fixed effects	Estimate	SE	*z*-value	Pr(>|*z*|)
(Intercept)	0.74	0.21	3.57	
Style	0.43	0.17	2.54	0.011 *
TalkerGroup1 (Native English vs. Native Mandarin-High)	0.74	0.42	1.76	0.078
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	2.01	0.41	4.95	<0.001***
Style: TalkerGroup1	0.74	0.43	1.71	0.087
Style: TalkerGroup2	1.01	0.43	2.34	0.019 *
Post-hoc Tukey test examining the effect of Style (plain vs clear)				
Talker Group	Estimate	SE	*z*-ratio	*p*-value
Native English	−0.8	0.28	−2.87	0.004**
Native Mandarin-High	−0.57	0.28	−2.01	0.045*
Native Mandarin-Low	0.07	0.26	0.27	0.79

A. Intelligibility improvement based on clear speech enhancements

The results of the intelligibility perception experiment demonstrate that non-native talkers' proficiency level impacts their ability to produce intelligible speech in general, as well as their ability to increase intelligibility of their speech. Specifically, the current results show not only that higher-proficiency talkers' speech was generally more intelligible than lower-proficiency talkers' speech, but also that higher-proficiency talkers' clear speech enhancements resulted in larger plain-to-clear speech intelligibility gains than those of lower-proficiency non-native talkers. These results suggest that as non-native talkers' target language proficiency develops, they produce more intelligible speech and are better able to further increase the intelligibility of their speech.

It is possible that the difference in the size of intelligibility improvement between higher- and lower-proficiency talkers' clear speech stems from the difference in the types of acoustic modifications made in their clear speech. Lower-proficiency talkers' acoustic-phonetic modifications made in clear speech were overall smaller than those made in higher-proficiency talkers and native English talkers' clear speech. Thus, it is plausible that small overall changes in acoustic characteristics between plain and clear speech resulted in small changes in intelligibility in perception.

Acoustic characteristics of clear speech enhancements may also explain some of the individual variability observed in intelligibility improvement. Though the clear speech enhancements produced by native English talkers and higher-proficiency non-native talkers resulted in larger intelligibility gains than those produced by lower-proficiency non-native talkers (at the noise levels used in the current study), there was a noteworthy degree of individual variability within each talker group. Particularly, among the native English talkers, Talker 405's clear speech resulted in smaller intelligibility gains (3% increase) compared to the other three native English talkers (15%, 17%, and 15% increase). Also, among the higher-proficiency non-native talkers, Talker 306's and 310's clear speech resulted in relatively large intelligibility gains (19% and 16% increase, respectively), though Talker 302's and 311's clear speech intelligibly gains were small (4% increase and 2% decrease, respectively). Similarly, lower-proficiency talkers exhibited variability in intelligibility gains (Talker 103, 3% decrease; Talker 104, 5% increase; Talker 106, no change; Talker 107, 12% decrease). The substantial individual variability in clear speech intelligibility improvement has also been reported in previous studies, including when native English listeners evaluated native English talkers' plain and clear speech (Ferguson, 2004; Smiljanić and Bradlow, 2005), as well as when native English listeners evaluated highly proficient non-native talkers' plain and clear speech (Smiljanić and Bradlow, 2011).

Given some previous results demonstrating that a larger degree of vowel space expansion is associated with a larger plain-to-clear speech intelligibility improvement (Ferguson, 2004; Ferguson and Kewley-Port, 2007), it is possible that the degree of vowel space expansion may explain some of the individual variability in the size of intelligibility gains. In fact, the native talker 405, who showed the least amount of intelligibility gains, also showed the smallest vowel space expansion among other native talkers. However, this is less clear for non-native talkers. For example, the higher-proficiency talker 306, who showed the smallest vowel space expansion, showed the largest intelligibility gains among other higher-proficiency talkers. Also, the lower-proficiency talker 107, who showed the largest vowel space expansion among other lower-proficiency talkers, showed 12% *decrease* in clear speech intelligibility. Thus, it is possible that the relationship between certain acoustic characteristics of clear speech enhancements and intelligibility gains may be different for native listeners' perception of native and non-native clear speech. Further, acoustic features not examined in the materials of the current perception study (e.g., the extent of coarticulation; Bradlow, 2002; frequency of stop-burst releases: Ferguson and Kewley-Port, 2007) may also explain the clear speech intelligibility benefit, in terms of the differences among different talker groups as well as the individual variability within each talker group.

Indeed, no single acoustic or articulatory modification investigated here correlated directly with intelligibility improvements in the present study. A number of factors could account for this. It is possible that we did not examine a key acoustic or articulatory feature that is driving the results. Alternately, it is possible, and we would argue, likely, that a constellation of features is jointly responsible for intelligibility gains. That is, because speech perception is typically done in a holistic way, it is likely that multiple features "conspire" together to result in intelligibility gains. In fact, it is possible that different features may impact perception differently for different talkers or groups of talkers. For example, vowel space expansion may result in improved intelligibility when listening to native English speakers, but perhaps modulations in speaking rate would be more beneficial when listening to a non-native speaker. A future study with a larger dataset may be better suited to explore the relationship between acoustic-phonetic variations and variability in clear speech intelligibility improvement, and how it may differ for the speech of native talkers and non-native talkers of different proficiency levels.

It is also important to note that the SNR used in the perception experiment is not equivalent for native vs non-native talker groups (i.e., −6 dB SNR for native talkers' speech and −2 dB SNR for non-native talkers' speech). Typically, given equivalent SNRs, non-native speech is less intelligible than native speech (Munro, 1998; Rogers *et al.*, 2004); therefore, changes in intelligibility from plain to clear speech may be subject to floor or ceiling effects if equal noise levels are used (e.g., Rogers *et al.*, 2010). While an investigation of intelligibility in quiet does not demonstrate differences between the native talkers and high-proficiency non-native talkers (see supplementary material[1]), we believe this is likely due to a ceiling effect. While we cannot strictly compare the clear speech intelligibility between native and non-native talker groups, at a minimum, we are able to directly compare the higher- and lower-proficiency talkers, whose speech is presented at the same SNR, and observe that proficiency does impact clear speech intelligibility benefits for these talkers.

It is possible that a higher noise level would differentially affect clear and plain speech for the different talker groups. Our goal in the present study was to equate the intelligibility of plain speech produced by all talker groups to the extent possible in order to investigate the magnitude of clear speech intelligibility benefits from that baseline level of performance. However, we believe that the differences in the clear speech intelligibility benefits that we observe between talker groups are unlikely to be due solely to the variance in noise levels across talker groups.

B. The influence of target language proficiency level on non-native clear speech enhancements

The acoustic analyses of the stimuli further revealed differences in the characteristics of native and non-native speech in general, as well as in their clear speech modifications. For example, in terms of the difference between native and non-native speech, non-native talkers (higher- and lower-proficiency) spoke more slowly than native English talkers in both plain- and clear-speaking styles. The overall speaking rate was similar between higher- and lower-proficiency talkers' speech. However, there was a clear difference between the plain-to-clear speech modifications made by higher- and lower-proficiency non-native talkers. That is, higher-proficiency talkers decreased their speaking rate in clear speech to a larger degree than lower-proficiency talkers did, suggesting that non-native talkers' overall speaking rate is partially independent from their ability to make plain-to-clear speech modifications in speaking rate. These findings suggest that non-native talkers' ability to make plain-to-clear speech modifications is influenced by their target-language proficiency level.

One outstanding question may concern why lower-proficiency talkers have difficulty varying acoustic characteristics of their speech between plain and clear speech. One possibility is that they have difficulty moving from hypospeech (Lindblom, 1990; operationalized as plain speech in the current study) to hyperspeech (operationalized as clear speech).

That is, while talkers may be able to minimize their articulatory effort in producing speech (hypospeech), they may have difficulty modifying phonetic characteristics to improve intelligibility for listeners (hyperspeech). This is illustrated in the plain-clear variations of the speaking rate for lower-proficiency talkers, where their clear-speech speaking rate did not slow down (i.e., did not move to the slower side within their range of speaking rate), as compared to native English and higher-proficiency non-native talkers.

However, the small range of plain-to-clear speech variations in lower-proficiency talkers' speech could also stem from their difficulty minimizing articulatory effort (hypospeech or plain speech in the current study). Though plain speech, in the current study and previous studies, is elicited by asking talkers to read materials in a laboratory setting (Smiljanić and Bradlow, 2009) and does not typically contain phonetic reductions that speech occurring in a more natural setting does, it is often produced with reduced effort compared to clear speech, as seen in the plain-to-clear speech variations in previous studies (e.g., Bradlow *et al.*, 2003; Picheny *et al.*, 1986). However, it is possible that lower-proficiency talkers were not necessarily speaking with reduced effort in plain speech, resulting in small plain-to-clear speech variations. In other words, lower-proficiency talkers may already have been exerting substantial effort to produce plain speech (and they may generally hyperarticulate to produce second language speech), thus there was not much room to 'enhance' in clear speech. This is illustrated in the small plain-to-clear variation in lower-proficiency talkers' vowel space originated from the relatively large vowel space in their plain speech; on the other hand, the large plain-to-clear variation in native English talkers' vowel space originated from relatively the small vowel space in their plain speech. This suggests that lower-proficiency talkers did not necessarily produce the plain speech with reduced effort compared to their clear speech. Together, the current results suggest that the size of plain-to-clear speech modifications could be influenced by the talker's ability to enhance acoustic characteristics in clear speech as well as to reduce articulatory effort in plain speech. This could be compatible with the previous work suggesting that non-native speech production is more effortful for talkers with lower proficiency (e.g., Kormos and Dénes, 2004; Nip and Blumenfeld, 2015; Poulisse, 1997). In conjunction with the current results, it is possible that as non-native talkers' proficiency level develops, their speech production generally becomes less effortful (e.g., in plain speech), and ultimately there is more room for talkers to enhance characteristics of their speech. Furthermore, the influence of non-native talkers' proficiency level on the size of plain-to-clear speech modifications suggests that knowing how to vary acoustic characteristics of speech (e.g., in plain and clear speech) may be a part of the skill set that contributes to the target language proficiency.

It is also possible that talkers in the lower-proficiency group might produce acoustic modifications that are not only smaller in magnitude (i.e., less differentiation between speaking rates and $F0$ mean and range in clear- and plain-speaking styles), but also are "wrong" in some other way. For example, the collection of acoustic modifications could be an ineffective ensemble of modifications to yield intelligibility gains in perception. Alternately, talkers could be changing acoustic characteristics, but not in the direction that native speakers alter their acoustic characteristics. This question is an open one, in terms of the perceptual consequences of the acoustic characteristics, in part because the perception tasks typically used in clear speech research allow for (and perhaps even encourage) use of a holistic set of cues. That is, it is likely that listeners use a constellation of acoustic features, and these features likely work in concert to create the intelligibility benefits associated with clear speech. However, it is difficult to holistically assess the acoustic manipulations talkers make. Nevertheless, future work, examining both acoustic details and their impact on perception, can continue to investigate the relationships between specific acoustic manipulations and intelligibility.

V. CONCLUSION

The current study examined intelligibility benefits resulting from native and non-native clear speech. Specifically, we asked whether clear speech enhancements produced by native and non-native talkers of different proficiency levels result in a similar intelligibility improvement for native English listeners. The results demonstrate that non-native talkers' target language proficiency level impacts their speech intelligibility and the enhancement of that intelligibility. That is, the higher-proficiency talkers' speech was generally more intelligible than lower-proficiency talkers' speech, and higher-proficiency talkers' clear speech resulted in a larger plain-to-clear speech intelligibility improvement than lower-proficiency talkers' clear speech. The results of the acoustic analysis demonstrated that the degrees of plain-to-clear speech modifications were much smaller for lower-proficiency non-native talkers' speech compared to those of higher-proficiency talkers' and native talkers' speech. The higher-proficiency talkers' acoustic modifications were comparable to those of native talkers. Together, these results suggest that non-native talkers' ability to acoustically enhance characteristics of their speech and increase intelligibility of their speech improves as their target language proficiency level develops. A future investigation could directly examine the relationship between specific acoustic manipulations and intelligibility improvement for different combinations of native and non-native talkers and listeners.

ACKNOWLEDGMENTS

This work was funded by National Science Foundation Grant Nos. BCS-194173 to M.M.B.B. and M.K., BCS-1734166 to M.M.B.B., as well as a Lokey Doctoral Science Fellowship, National Federation of Modern Language Teachers Associations Dissertation Support Grant, and Institute of Cognitive and Decision Sciences Dissertation Research Award to M.K. We would like to thank Kaylynn Gunter and Charlie Farrington for their help with the acoustic analysis of the production data. We thank Aubrey Whitty for her assistance with the analysis of the perception data.

APPENDIX

List of 30 BKB sentences recorded by native and non-native English talkers. Keywords used for intelligibility scoring are underlined.

Sentence
The shop closed for lunch.
Some nice people are coming.
They met some friends.
Flowers grow in the garden.
The train stops at the station.
The puppy plays with a ball.
Mother cut the birthday cake.
He closed his eyes.
The raincoat is very wet.
She is paying for her bread.
Some men shave in the morning.
The driver lost his way.
The oven door was open.
The car is going too fast.
The silly boy is hiding.
The apple pie is baking.
The sky was very blue.
People are going home.
She is calling for her daughter.
He is skating with his friend.
They painted the wall.
The dog is eating some meat.
A boy broke the fence.
The snow is on the roof.
The bath water was warm.
He is reaching for his spoon.
The boy got into trouble.
He paid his bill.
Mother made some curtains.
The man tied his shoes.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0009403 for additional information; for a description of the comparison of the noise and quiet data; a figure illustrating this comparison; and a statistical analysis comparing the two conditions.

[2]The talker group and speaking style impacted the speaking rate (with pauses) and articulation rate (without pauses) in a similar manner; thus, the current paper includes the results of the speaking rate.

[3]We used ANOVA instead of mixed-effects regression models in order to avoid the risk of overfitting (e.g., Crawley, 2002), as there are only four talkers in each talker group.

[4]Following Lakens (2013), we report two types of effect size statistics: partial eta-squared (η_p^2) and generalized eta-squared (η_G^2). Generalized eta-squared (η_G^2) is a recommended effect size statistic for repeated measure designs (used for later analyses in this paper; Bakeman, 2005).

[5]It is important to note that these stimuli were chosen for use in a perception study, and not, specifically, to investigate vowel space expansion in clear speech. We used speech materials that were sufficiently easy for lower-proficiency talkers to read aloud, but these stimuli were not chosen because of the frequency of any specific vowel class.

[6]Here, we embed the speech with noise, following previous work (Bradlow and Alexander, 2007; Bradlow and Bent, 2002; Kato and Baese-Berk, in press; Rogers *et al.*, 2010; Smiljanić and Bradlow, 2011). In the supplementary material,[1] we also provide information on the transcription performance when participants listened to the target stimuli in quiet and compare performance in the noise and quiet conditions.

[7]Listeners' keyword recognition accuracy in the noise condition was lower than the accuracy in the quiet condition for both Native English and Native Mandarin (High and Low) talker groups' speech, and this is confirmed by the statistical analysis; see supplementary materials.[1] Thus, here, we focus on the data in the noise condition in order to examine whether clear speech enhancements improved listeners' keyword recognition in the challenging listening condition.

[8]It should be noted that our experimental design confounds signal-to-noise ratio and talker group for high-proficiency talkers and native English talkers. In spite of this experimental confound, we believe our results suggest that language proficiency may play a role in clear speech enhancements.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**(3), 379–384.

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

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer," (computer program), http://www.praat.org/ (Last viewed November 8, 2021).

Bradlow, A. R. (**2002**). "Confluent talker- and listener-related forces in clear speech production," in *Laboratory Phonology*, Vol. VII, edited by C. Gussenhoven and N. Warner. (Mouton de Gruyter, Berlin), pp. 241–273.

Bradlow, A. R., and Alexander, J. (**2007**). "Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**(4), 2339–2349.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**(1), 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Crawley, M. J. (**2002**). *Statistical Computing: An Introduction to Data Analysis Using S-Plus* (John Wiley & Sons Ltd, London).

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**(4), 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. **50**, 1241–1255.

Ferguson, S. H., and Morgan, S. D. (**2018**). "Talker differences in clear and conversational speech: Perceived sentence clarity for young adults with normal hearing and older adults with hearing loss," J. Speech Lang. Hear. Res. **61**(1), 159–173.

Gagné, J. P., Masterson, V., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). "Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech," J. Acad. Rehab. Aud. **27**, 135–158.

Gagné, J. P., Rochette, A. J., and Charest, M. (**2002**). "Auditory, visual and audiovisual clear speech," Speech Commun. **37**(3-4), 213–230.

Gerstman, L. (**1968**). "Classification of self-normalized vowels," IEEE Trans. Audio Electroacoustics **16**(1), 78–80.

Granlund, S., Hazan, V., and Baker, R. (**2012**). "An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals," J. Phon. **40**(3), 509–520.

Imai, S., Walley, A. C., and Flege, J. E. (**2005**). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.

Kallay, J. E., and Redford, M. A. (**2018**). "Coarticulatory effects on 'the' production in child and adult speech," in Proceedings of the 9th International Conference on Speech Prosody, Poznan, Poland, pp. 1004–1007.

Kato, M., and Baese-Berk, M. M. (**2021**). "Contextually-relevant enhancement of non-native phonetic contrasts," J. Phon. **88**, 101099.
Kormos, J., and Dénes, M. (**2004**). "Exploring measures and perceptions of fluency in the speech of second language learners," Systems **32**, 145–164.
Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**(5), 2165–2172.
Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.
Lakens, D. (**2013**). "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," Front. Psych. **4**, 1–12.
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Amsterdam), pp. 403–439.
Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (**2004**). "Clear speech perception in acoustic and electrical hearing," J. Acoust. Soc. Am. **116**, 2374–2383.
McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (**2017**). "Montreal Forced Aligner" (computer program), Version 0.9.0, http://montrealcorpustools.github.io/Montreal-Forced-Aligner/ (Last viewed November 8, 2021).
McCloy, D. R. (**2016**). "phonR: tools for phoneticians and phonologists," R package version 1.0-7.
Munro, M. J. (**1998**). "The effects of noise on the intelligibility of foreign-accented speech," Stud. Second Lang. Acq. **20**(2), 139–154.
Nearey, T. M. (**1977**). "Phonetic feature systems for vowels," Ph.D. dissertation, University of Connecticut.
Nip, I. S., and Blumenfeld, H. K. (**2015**). "Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners," J. Speech Lang. Hear. Res. **58**(3), 653–668.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Lang. Hear. Res. **28**(1), 96–103.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.
Poulisse, N. (**1997**). "Language production in bilinguals," in *Tutorials in Bilingualism: Psycholinguistic Perspectives*, edited by A. M. B. de Groot and J. F. Kroll (Erlbaum, Mahwah, NJ), pp. 201–224.
Rogers, C. L., Dalby, J., and Nishi, K. (**2004**). "Effects of noise and proficiency on intelligibility of Chinese-accented English," Lang. Speech **47**(2), 139–154.
Rogers, C. L., DeMasi, T. M., and Krause, J. C. (**2010**). "Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers," J. Acoust. Soc. Am. **128**(1), 410–423.
Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gormon, K., Prichard, H., and Yuan, J. (**2014**). FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2, available from https://doi.org/10.5281/zenodo.22281 (Last viewed November 8, 2021).
Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol. **7**, 212–218.
Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (**2020**). "afex: Analysis of Factorial Experiments. R package version 0.27-2," https://CRAN.R-project.org/package=afex (Last viewed March 14, 2020).
Smiljanić, R. (**2021**). "Clear speech perception," in *The Handbook of Speech Perception*, 2nd ed., edited by L. C. Nygaard, J. Pardo, D. B. Pisoni, and R. Remez (Wiley Publishing, New York), pp. 177–205.
Smiljanić, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**(3), 1677–1688.
Smiljanić, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Ling. Compass **3**(1), 236–264.
Smiljanić, R., and Bradlow, A. R. (**2011**). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," J. Acoust. Soc. Am **130**(6), 4020–4031.
Thomas, and E. R., Kendall, T. (**2015**). "NORM: The vowel normalization and plotting suite," http://lingtools.uoregon.edu/norm/norm1 (Last viewed November 8, 2021).
Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Lang. Hear. Res. **39**(3), 494–509.