Journal of Phonetics 88 (2021) 101099

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Contextually-relevant enhancement of non-native phonetic contrasts

Misaki Kato *, Melissa M. Baese-Berk

Department of Linguistics, University of Oregon, Eugene, OR 97403, United States

ARTICLE INFO

Article history:
Received 17 December 2020
Received in revised form 2 July 2021
Accepted 23 August 2021
Available online 14 September 2021

Keywords:
Speech production
Second language acquisition
Speech enhancement

ABSTRACT

One important factor that contributes to successful speech communication is an individual's ability to speak more clearly when their listeners have difficulty understanding their speech. Though previous studies have demonstrated that native talkers implement acoustic–phonetic speech enhancements to ensure that their speech is understood by listeners, how non-native talkers employ goal-oriented enhancements is less well-understood. Here, we examine acoustic characteristics of speech enhancements produced by native and non-native English talkers of varying proficiency. Specifically, we investigate native Mandarin learners of English. The results show that non-native talkers' ability to enhance a specific sound contrast differed depending on their familiarity with the target English contrast from their native language experience (Mandarin), as well as their English proficiency level. These results highlight that talkers are able to enhance their speech in native and non-native languages, but also suggest that this flexibility is shaped by the talkers' target language proficiency and the type of acoustic manipulation involved in the adaptation.

1. Introduction

One important factor that contributes to successful speech communication is an individual's ability to speak more clearly when their listeners do not understand their speech. It has been widely demonstrated that talkers are able to enhance various features of their speech (e.g., by speaking more slowly, loudly, or by articulating sounds more clearly) to make their speech more understandable to their listeners (e.g., Picheny et al., 1986). While speech enhancement strategies used by native talkers have been examined in a variety of communication contexts (e.g., Hazan & Baker, 2011; Scarborough & Zellou, 2013), our understanding of non-native talkers' speech enhancement strategies is mostly limited to those examined in a context where they read materials as if talking to a hearing-impaired listener (e.g., Smiljanić & Bradlow, 2011). Particularly, while previous work has demonstrated that native talkers make targeted acoustic modifications to enhance characteristics of particular sound contrasts in a communicative task (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Seyfarth et al., 2016), it is unknown how such targeted enhancements are implemented by non-native talkers of different proficiency levels. Thus, the current study examines native and non-native English talkers' ability to produce speech enhancements. Specifically, we examine talkers' speech enhancements using a word-reading paradigm following Baese-Berk & Goldrick (2009), where talkers communicate target words (e.g., *cap*) to a listener when a phonetically similar minimal-pair neighbor (e.g., *cab*) either is or is not present in the context. We examine acoustic characteristics of speech modifications made in such contexts, asking how talkers' native language status and non-native talkers' proficiency level impact the size of these modifications. For example, do talkers lengthen voice onset time of the initial stop in words like *peer* when a minimal pair competitor (e.g., *beer*) is presented simultaneously more than when such a minimal pair word is not present? Are native and non-native talkers equally likely to make this enhancement? Furthermore, we ask how the effects of talkers' target language experience on the contextually-relevant enhancements differ depending on the talkers' familiarity with the target sound contrast (i.e., a contrast that also exists in non-native talkers' native language vs. a contrast that does not). For example, do native Mandarin speakers of English, who use similar voicing contrasts word-initially in English and Mandarin, enhance contrasts word-initially, but not word-finally, where Mandarin does not have such contrasts?

* Corresponding author.
E-mail addresses: misaki@uoregon.edu (M. Kato), mbaesebe@uoregon.edu (M.M. Baese-Berk).

https://doi.org/10.1016/j.wocn.2021.101099
0095-4470/

1.1. Speech enhancements in different tasks from different talkers

Previous studies have demonstrated that talkers are able to enhance various acoustic–phonetic features of their speech to make it more intelligible to their listeners (e.g., Uchanski, 2005). One way to examine talkers' speech enhancement strategies is to investigate clear speech, a speaking style that talkers use when they are aware that the listeners may have difficulty understanding them (e.g., Smiljanić & Bradlow, 2009; Uchanski, 2005). Previous studies have shown that in clear speech, native talkers of the language use a range of acoustic–phonetic modifications and these modifications result in robust intelligibility gains for listeners of various characteristics, including hearing-impaired listeners as well as non-native listeners (e.g., Bradlow & Bent, 2002; Liu et al., 2004; Picheny et al., 1985; Schum, 1996).

However, previous studies have demonstrated that speech enhancements are not uniform phenomena (Tuomainen & Hazan, 2018). That is, the characteristics of speech enhancements differ depending on what type of task they engage in to produce the enhancements. For example, native talkers' speech enhancements elicited in read speech involved more extreme changes in some acoustic–phonetic characteristics (e.g., pitch range, speaking rate, vowel duration, vowel space) than speech enhancements elicited in spontaneous speech (e.g., Hazan & Baker, 2011; Scarborough & Zellou, 2013). The presence of a listener also impacts talkers' acoustic modifications; when producing foreigner-directed speech, native talkers employ more extreme changes in durations and vowel space, when talking to an imagined non-native listener compared to when talking to a real non-native listener present in the room (Scarborough et al., 2007). These results suggest that speech enhancements must be investigated in different contexts.

Despite the wealth of information on speech enhancement strategies, these studies have focused on speech produced by native talkers, and investigations of strategies that are employed by non-native talkers are limited. Those studies which have investigated clear speech from non-native talkers have suggested that proficiency in the target language may impact what speech enhancement strategies the talkers use. For example, highly proficient non-native talkers make clear speech adjustments that are similar to those made by native talkers in terms of modifications of vowel space, F0, intensity, and temporal characteristics (e.g., word duration, articulation rate of sentences; Bradlow, 2002; Granlund et al., 2012). However, the size of plain-to-clear speech modifications made by earlier second language (L2) learners is much smaller compared to those made by later L2 learners (Rogers et al., 2010).

Other studies have suggested that such an effect of proficiency level on non-native speech enhancements may differ depending on the focus of the acoustic modifications. For example, native Croatian talkers manipulated vowel duration to a larger extent in Croatian clear speech than native English talkers did in English clear speech, reflecting the difference in the importance of duration cues between Croatian and English (Smiljanic & Bradlow, 2008). Such differences in the use of acoustic cues in different languages could influence how easy or difficult it is to make certain segmental enhancements for non-native talkers of the language and this may also differ depending on the proficiency of talkers in their L2.

The difficulty of manipulating acoustic cues in a non-native language could be impacted by non-native talkers' experience with their native language. It has been widely documented that L2 learners' native language influences their learning of L2 (e.g., Lado, 1957, Flege, 1995), and that L2 sounds that exist in learners' native language are easier to learn to produce compared to L2 sounds that do not (e.g., Brière, 1966; Vokic, 2008). Thus, it is possible that such ease and difficulty associated with non-native sound production extends to enhancements of non-native segments. That is, making segmental enhancements could be more challenging for non-native sounds that do not exist in the talker's native language than the non-native sounds that do, especially for inexperienced non-native talkers. Indeed, previous work has suggested that late English learners' segmental modifications led to a decreased intelligibility for an English vowel that does not exist in their native language (e.g., /ɪ/ in *bid* for native Spanish speakers producing English vowels; Rogers et al., 2010). However, proficient non-native talkers are able to make segmental modifications to enhance non-native contrasts that do not exist in their native language (e.g., /ɛ/ vs. /æ/ for Korean learners of English; Hwang et al., 2015). These results suggest that making non-native acoustic adjustments that talkers are not used to making in their native language can be generally more difficult than those that they are familiar with from their native language experience; though more experienced, higher-proficiency talkers are able to enhance non-native contrasts that exist in their native language as well as those that do not.

Taken together, previous studies have demonstrated that production of speech enhancements can be affected by multiple factors. Specifically, for non-native talkers, their target language proficiency level influences the quality of clear speech modifications when explicitly asked to read materials clearly. Furthermore, talkers' experience with the target language sound system (e.g., native vs. non-native status, non-native talkers' proficiency level) may impact speech enhancements differently depending on the focus of the acoustic enhancements. Previous studies also demonstrate that the characteristics of speech enhancements differ depending on the task, though this has only been investigated in native talkers' speech. Thus, it is not clear how factors such as non-native talkers' L2 proficiency level and focus of acoustic enhancements impact talkers' L2 speech enhancements in a more ecologically valid communication context (i.e., when talkers are not given explicit instruction to speak clearly).

A further question is how learners acquire contrasts more broadly. Many previous studies have examined L2 contrasts that do not exist anywhere in the learner's native language. However, it may be the case that learners have familiarity with a contrast in one position (e.g., word-initially) but not in another (e.g., word-finally). One open question is whether the learners acquires one contrast which they use regardless of position, or whether they acquire a more position-specific notion of contrast. This is critically important because contrasts are not always realized the same way across phonotactic positions. For example, stop voicing in English word-initially is typically signaled by voice onset time (shorter for /b/ than for /p/);

however, word-finally, the primary cue to stop voicing is duration of the previous vowel (longer for /b/ than for /p/). Therefore, learners must acquire position-specific cues to voicing to accurately produce this contrast across contexts. However, relatively little work has directly investigated how these contrasts are acquired or enhanced.

In the current study, we examine how native and non-native talkers make speech enhancements in a task similar to a naturalistic talker-listener interaction. Specifically, we examine how these talkers accommodate their speech when the potential communication difficulty is signaled in the context implicitly, rather than when it is signaled by explicit instructions to read materials clearly.

1.2. Contextually-relevant speech enhancements

Previous work has demonstrated that talkers make contextually-relevant speech enhancements. Specifically, when a listener misunderstands a particular part of an utterance (e.g., a specific word), talkers selectively enhance that part of the utterance to correct the misunderstanding (e.g., Maniwa et al., 2009; Ohala, 1994; Oviatt et al., 1998; Schertz, 2013; Stent et al., 2008). For example, when native English talkers spoke to a simulated listener and received feedback that the utterance was misunderstood (e.g., the talker says "pit" but the simulated partner guesses "bit"), the talkers enhanced the misunderstood contrast by manipulating a relevant acoustic feature (e.g., VOTs of the /p/ and /b/) on a second repetition (Schertz, 2013). This type of targeted error correction did not occur when the talker received an open-ended request for repetition (e.g., "???").

Talkers make contextually-relevant speech enhancements in a communicative task even without feedback from the listener. For example, in a communicative task requiring a talker to convey information to a listener, native English talkers exaggerated differences in VOTs of English word-initial consonants (e.g., /p/-/b/) when a target word to communicate (e.g., *pill*) was displayed with a minimal pair word (e.g., *bill*), compared to when it was presented only with unrelated words (Baese-Berk & Goldrick, 2009; Buz et al., 2014, 2016; Seyfarth et al., 2016). Though the investigation of such contextually-relevant hyperarticulation has mostly been limited to native talkers' productions, one study demonstrated that highly proficient non-native talkers exaggerated a non-native contrast (e.g., /æ/-/ɛ/) when a target word (e.g., *sat*) displayed with a similar word (e.g., *set*) in a word-communication task (Hwang et al. 2015). However, it is unknown how such contextually-relevant speech enhancements are made by non-native talkers of differing target language proficiency levels.

1.3. Current study

In the current study, we examine acoustic characteristics of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency. We use a word-reading paradigm that has been shown to elicit contextually-relevant speech enhancements (e.g., Baese-Berk & Goldrick, 2009). We ask whether the effect of talkers' language experience differs depending on the type of acoustic enhancements examined; specifically, enhancements of consonants in a sound contrast that exists in talkers' native and non-native languages (henceforth L1L2 contrast) and consonants in a sound contrast that does not exist in the talker's native language but does exist in the talker's L2 (henceforth L2-only contrast).

2. Methods

2.1. Participants

Thirty-four native English talkers (29 females, 5 males; age 18–22, mean = 19) and 44 native Mandarin talkers (34 females, 10 males; age = 19–35, mean = 24.7) participated in the study.[1] Native English talkers were recruited from the Linguistics and Psychology Human Subject Pool at the public university in the Pacific Northwest, and were given partial course credit for their participation. Native Mandarin participants were either paid or given partial course credit for their participation.

The non-native English talkers were classified into lower- and higher-proficiency talkers, based on their most recent English proficiency test score. Talkers who had reported a Test of English as a Foreign Language (TOEFL) score of lower than 72[2] were categorized as lower-proficiency native Mandarin (Native Mandarin-Low) talkers (n = 22). The talkers who had reported a TOEFL score of higher than 72 were categorized as higher-proficiency native Mandarin (Native Mandarin-High) talkers (n = 22).[3] Online Supplementary Material provides information regarding non-native (native Mandarin) talkers' English learning background and proficiency.

2.2. Materials and procedure

Participants first completed a context-production task, followed by a sentence-reading task. In the context-production task, target words were 80 English monosyllabic words. Forty targets consisted of 20 minimal pairs that contrasted consonants and the other 40 targets consisted of 20 minimal pairs that contrasted vowels. Only the consonant production data is reported in the current manuscript (see Supplementary Material for the list of consonant target words). The 20 consonant targets (i.e., 10 minimal pairs) contained a phonemic contrast that exists in both L1 Mandarin and L2 English for the native Mandarin talkers (L1L2 consonant targets) and 20 targets (i.e., 10 minimal pairs) contained a phonemic contrast that exists only in L2 English and not L1 Mandarin for the native Mandarin talkers (L2-only consonant targets). The 20 L1L2 consonant targets contrasted /p/ and /b/ in word-initial position

[1] The stopping rule for our data collection was to collect data until the end of the term, or until we had 22 participants (to match the number of participants in each of the Mandarin talker groups described below). Due to restrictions within our subject pool, we often collect participant data we cannot use because, for example, participants are not native English speakers or report a history of speech or hearing impairment. Therefore, we often collect data from a larger number of participants than we minimally need from the Human Subject Pool. At the end of the term, we had data from 34 native English talkers, so this was the final sample size reported here.

[2] The classification was done based on the proficiency level classification provided by TOEFL. https://www.ets.org/toefl/institutions/scores/interpret/

[3] 7 talkers did not report their TOEFL score, but reported their International English Language Testing System (IELTS) score; for these talkers, their IELTS score was converted to a TOEFL score based on the conversion table provided in Educational Testing Service (2010). When a talker provided neither their TOEFL nor IELTS score (n = 4), their perceived accentedness score, collected for another study using the same stimuli (see supplementary material) was used as a proxy for their proficiency level.

(e.g., *peer* vs. *beer*). The 20 L2-only consonant targets contrasted /p/ and /b/ in word-final position (e.g., *cap vs. cab*). Stimuli were designed such that the target items were matched for vowels across conditions, so that qualities of the stimuli unrelated to the target consonants would be less likely to impact our results. The 20 fillers were English monosyllabic words that did not contain the target contrasts.

The context-production task was modeled after the word-reading paradigm used in Baese-Berk and Goldrick (2009) and Buz et al. (2016), and was administered using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). A simulated partner paradigm was used, where the participant was told that they would interact with a partner online, but a computer actually provided responses (Buz et al., 2016). In order to familiarize the participant with the role that their partner would later play in the context-production task, the task began with five perception trials, where the participant saw three words on the screen and heard a male native English speaker say, *"Click on the ___ now."* The participant was instructed to choose one of the three words that the speaker said as quickly and accurately as possible. The five perception trials consisted of 2 trials that had a consonant or a vowel minimal pair (i.e., *star, loom,* ***room****; sat,* ***set****, oil*), which were not the target contrasts in the context-production task. The other 3 perception trials did not have minimal pairs. The 2 trials with minimal pairs were included in the perception trials in order to familiarize the participant with the potential difficulty in choosing the correct word that their partner might experience in the following part of the context-production task.

After the perception trials, the participants were told that they would now give instructions to their partner. Participants saw a message that the computer was searching for their partner online. After a few seconds, participants were told that they had been matched with a partner online and proceeded to the production part of the task.

On each trial, the participant was presented with three words on the screen. Then, one of the three words was highlighted, and the participant was asked to produce the highlighted word (i.e., the target) in the phrase, *"Click on the TARGET now"*, for a partner, who could also see the three words but did not know which of the three was the target. After a delay (i.e., 800ms, 1200ms, 2000ms, 4000ms, 6000ms; randomly assigned for items), the participant was informed that their partner made a response but was not informed which word the partner selected, then the trial advanced. In Context conditions, both the target and its minimal pair neighbor were presented on the screen with a filler as the third word (e.g., ***peer,*** *beer, town*). In No Context conditions, the target was presented with two fillers (e.g., *soft,* ***peer****, noon*). In Filler trials, three fillers were presented.

After three practice trials, each participant completed 60 test trials. Of the 60 test trials, 20 trials were with consonant targets (i.e., 10 trials each with L1L2 consonant targets and L2-only consonant targets); five trials of L1L2 consonant targets and five trials of L2-only consonant targets were presented in Context conditions and the other half were presented in No Context conditions. The rest of the test trials (i.e., 40 trials) were trials with vowel targets (20 trials; analyses not included in the current paper), and filler trials (20 trials). In order to ensure that one participant did not produce both targets in a minimal pair (e.g., *pad*, *bad*), participants were divided into two groups. The consonant voicing was counterbalanced across participants, such that a participant in one group produced an L1L2 consonant target *pad* in the Context condition, and another participant in a different group produced its minimal pair *bad* in the Context condition. Following the context-production task, participants completed a sentence reading task which was used for the accentedness judgment task (see Supplementary Material).

2.3. Acoustic analysis

In the context-production task, one talker produced a total of 20 consonant targets and 20 vowel targets. That is, for consonant trials, each talker produced 10 L1L2 consonant targets (/p/-/b/ in word initial position; five targets in Context condition and five targets in No Context condition) and 10 L2-only consonant targets (/p/-/b/ in word final position; five targets in Context condition and five targets in No Context condition). Thus, there was a total of 3120 items (78 talkers $\times$ 40 targets). Productions with mispronunciation and disfluency (e.g., repetition) were excluded from the acoustic analysis. The excluded items were 46 out of 3120 items (i.e., less than 1% of the total number of items). Thus, for the consonant targets presented here, we analyzed a total of 1541 items: 773 L1L2 consonant targets, 768 L2-only consonant targets. Praat (Boersma & Weenink, 2001) was used for all measurements of the acoustic analysis.

We examined several acoustic features for L1L2 consonant targets and L2-only consonant targets (see Supplementary Material for more detailed explanation of the acoustic analyses conducted here). For L1L2 consonant targets (i.e., /p/- and /b/-initial words), the voice onset times (VOTs) of /p/ and /b/ were manually annotated and measured. VOT was measured from the beginning of the stop burst on the waveform to the onset of the following vowel, which was defined as the left zero crossing of the first complete periodic cycle (Baese-Berk & Goldrick, 2009). We normalized this measure for whole word duration – the raw VOT was divided by the duration of the whole word, as a proxy for speaking rate. For L2-only consonant targets (i.e., /p/- and /b/-final words), we annotated and measured the durations of the vowels preceding the target consonants, which we normalized for whole word duration. We also used voicing proportions of the target consonants (i.e., C2: word-final /p/ and /b/) as another measure to examine production patterns of the coda voicing contrast. To measure C2 voicing proportions, we used Praat to count the total number of voiced 10 ms frames in each C2 (Seyfarth et al., 2016). Of the 768 L2-only consonant contrast items, we found that 26 items were not released (i.e., about 3%). Because we could not reliably measure the durations of the target consonant closure or burst without the release, we excluded these 26 items from the analyses of normalized vowel durations and C2 voicing proportions. All the annotations were made by the first author. A research assistant annotated 5% of the L1L2 consonant targets and 5% of the L2-only consonant targets. For L1L2 consonant targets, the inter-rater agreement was calculated for VOT (r = 0.99, p < 0.001) and word duration (r = 0.99, p < 0.001). For L2-only consonant targets, the inter-rater agreement was calculated for vowel duration

(r = 0.95, p < 0.001), C2 duration (r = 0.95, p < 0.001), and word duration (r = 0.98, p < 0.001).

3. Results

In order to characterize segmental features of the target consonant contrasts, we analyzed several acoustic features, described above. For each acoustic feature (e.g., normalized VOTs for the /p/-/b/ contrast in word-initial position), we were interested in examining two questions. We first examined whether talkers in different Language Background groups (i.e., Native English, Native Mandarin-High, Native Mandarin-Low) generally differentiated between two phonemes (/p/ vs. /b/). However, our primary goal was to examine whether talkers manipulated acoustic features in one phoneme or another in order to enhance the contrast in the Context conditions as compared to No Context conditions. Previous results have demonstrated that talkers' speech enhancement patterns differ for different phonemes in a sound contrast (e.g., Seyfarth et al., 2016). Thus, we examined the size of talkers' phonetic adjustments in each phoneme and whether it differed for the productions of talkers in different Language Background groups.

Therefore, for each acoustic feature, we first present the results of a linear mixed-effects regression model, examining whether talkers made differences in the acoustic measure to distinguish one phoneme (e.g., /p/) from another (e.g., /b/), and whether the size of this difference was larger for one Language Background group's production than for another (e.g., Native Mandarin-High vs. Native Mandarin-Low), using the production data of both /p/- and /b/-targets in both Context and No Context conditions. Next, we examine production patterns of the two phonemes separately in order to explore how the difference in talkers' Language Background influenced production patterns in each phoneme (e.g., VOTs may differ for Native Mandarin-High vs. Low talkers' productions for /p/ but not for /b/), as well as how the difference in Context conditions (No Context vs. Context) influenced production patterns of each phoneme (e.g., VOTs may differ for No Context vs. Context productions for /p/ but not for /b/). Here, we also examine whether the effect of Context differed for the productions of talkers in different Language Background groups. By analyzing production patterns separately for each phoneme, we are able to examine the effects of talkers' Language Background and Context on production patterns of the target consonant contrasts closely, exploring *how* talkers in different groups make or enhance the contrast.

In the linear mixed-effects regression models presented below, fixed effects were Phoneme (/b/, /p/), Context (Context, No Context), and Language Background group (Native English, Native Mandarin-High, and Native Mandarin-Low); different combinations of these fixed effects were included in different models (see the description of each model below). Phoneme was contrast coded to compare between /b/-targets (0.5) and /p/-targets (-0.5). Context condition was contrast coded to compare between Context (0.5) and No Context (-0.5) conditions. Language Background was contrast coded to compare between Native English and Native Mandarin-High talkers (0.5, -0.5, 0) and between Native Mandarin-High and Native Mandarin-Low talkers (0, 0.5, -0.5)[4]. Models also included the maximal random effects structure that would converge, which included random intercepts for talker and item. The random effects structure also included a by-talker random slope for Phoneme or Condition (different depending on the model) and a by-item intercept for Language Background group. In each model, the random effects that did not account for significant amounts of variance were removed to avoid overfitting. *P*-values were calculated based on Satterthwaite approximations (Luke, 2017), using the lmerTest package for R (Kuznetsova, Brockoff, & Christensen, 2016). When explaining the results of the models, we only interpret the aspects that are relevant to the questions asked in each analysis; see Supplementary Material for model summaries which include the full results of each model.

3.1. L1L2 consonant contrast (/p/-/b/ in word-initial position)

Fig. 1 shows the mean normalized VOTs by Phoneme (/b/, /p/), Language Background (Native English, Native Mandarin-High, Native Mandarin-Low), and Context condition (No context, Context). The figure suggests that normalized VOTs were longer for /p/ than for /b/ for all Language Background groups' productions. However, this difference in the normalized VOTs seems to be larger for Native Mandarin talkers' (both High and Low) productions than for Native English talkers' productions; mean normalized VOT for /p/ – that for /b/ was 0.18 (Native English), 0.21 (Native Mandarin-High), 0.21 (Native Mandarin-Low). The linear mixed-effects regression model included Phoneme, Language Background, as well as interactions between the two factors as fixed effects (see Supplementary Material for the model syntax and summary of the results). There was a significant effect of Phoneme (β = −0.2, t = −38.8, p < 0.001). The effect of Phoneme interacted with the Native English vs. Native Mandarin-High comparison (β = 0.03, t = 2.48, p < 0.05), but not with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = 0.01, t = 0.79, p = 0.43). This indicates that the difference in normalized VOTs between /b/ and /p/ was larger for Native Mandarin-High talkers' productions than for Native English talkers' productions, but the difference was similar for Native Mandarin-High and Native Mandarin-Low talkers' productions. A post-hoc Tukey test confirmed that the effect of Phoneme was significant for all the Language Background group's productions: Native English (β = 0.18, SE = 0.008, t = 23.52, p < 0.0001), Native Mandarin-High (β = 0.21, SE = 0.01, t = 21.74, p < 0.0001), and Native Mandarin-Low (β = 0.21, SE = 0.01, t = 21.24, p < 0.0001).

Fig. 1 also suggests different patterns for normalized VOTs for each consonant (/b/ and /p/). For /b/, talkers did not change VOTs in Context conditions compared to No Context conditions. However, for /p/, normalized VOTs are influenced by both Language Background and Context condition. Further, talkers in all three Language Background groups increased VOTs from No Context to Context conditions. The linear mixed-effects regression models, used to analyze normalized

[4] Note that we include two pairwise comparisons among this three-level factor. The first pairwise comparison does not include the Native Mandarin-Low group (hence the "0″ in the contrast coding) and the second pairwise comparison does not include the Native English group (hence the "0" in the contrast coding for this comparison).

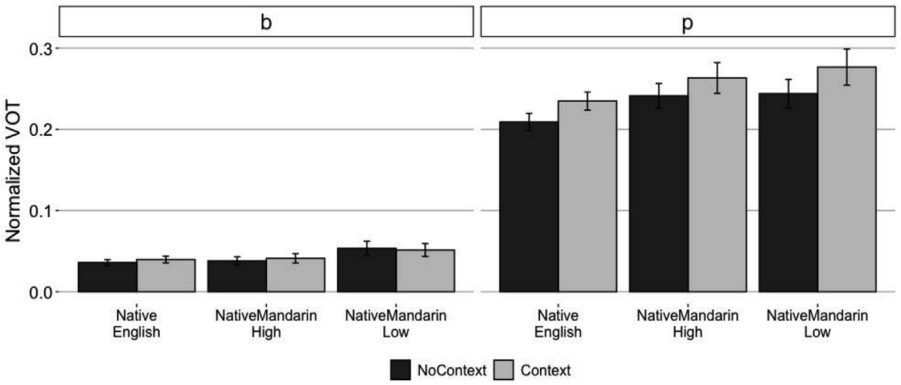

Fig. 1. Mean VOTs of the word-initial target consonants by Phoneme (/b/, /p/), Language Background group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

VOTs as the dependent variable separately for /b/-targets and /p/-targets, included Context condition (No Context, Context), Language Background, as well as interactions between the two as fixed effects (see Supplemental Material for the model syntax and summary of the results). For the /b/-target model, there were significant effects of Language Background: Native English vs. Native Mandarin-High (β = −0.01, t = −2.01, p < 0.05), Native Mandarin-High vs. Native Mandarin-Low (β = −0.02, t = −3.03, p < 0.01). This indicates that normalized VOTs for /b/ were longer for Native Mandarin-Low talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Context condition was not significant (β = 0.002, t = 1.31, p = 0.19). This pattern was similar for different Language Background groups' productions, as the effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = 0.001, t = 0.32, p = 0.75), or with the Native Mandarin-High and Native Mandarin-Low comparison (β = 0.003, t = 0.64, p = 0.52).

For the /p/-target model, there were significant effects of Language Background groups: Native English vs. Native Mandarin-High (β = −0.05, t = −3.29, p < 0.01), Native Mandarin-High vs. Native Mandarin-Low (β = −0.03, t = −2, p < 0.05). This indicates that normalized VOTs for /p/ were longer for Native Mandarin-Low talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Context condition was significant (β = 0.02, t = 4.49, p < 0.001). This pattern was similar for different Language Background groups' productions, as the effect of Context condition did not interact with the Native English vs. Native Mandarin-High comparison (β = 0.0006, t = 0.048, p = 0.96), or with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = −0.007, t = −0.52, p = 0.6).

Together, these results demonstrated that all talkers distinguished word-initial /b/ and /p/ with VOTs; normalized VOTs were longer for /p/ than for /b/. This difference was larger for Native Mandarin talkers' productions than for Native English talkers' productions (i.e., the VOT differences between /p/ and /b/ were larger for Native Mandarin-High than for Native English talkers, but were similar between the two Mandarin groups). This result is consistent with previous studies that have demonstrated that the difference between short- and long-lag stops is larger for Mandarin than for English (e.g., Chao & Chen, 2008; Rochet & Yanmei, 1991), suggesting an influence of the talkers' L1 on their differentiation between these two stops. Further, talkers manipulated VOTs in different types of Context conditions (No Context vs. Context) only for /p/; they increased VOTs for /p/ from No Context to Context conditions. The size of this Context condition-based difference for /p/ was similar for all Language Background group's productions. Thus, both native and non-native English talkers enhanced the word-initial /p/-/b/ contrast by increasing VOTs of /p/. These results are consistent with the previous work demonstrating that native talkers enhance word-initial plosive contrasts by elongating VOTs of the voiceless plosive (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016).

3.2. L2-only consonant contrast (/p/-/b/ in word final position)

3.2.1. Normalized vowel duration

The left panel in Fig. 2 shows the mean normalized durations of the vowels preceding the target consonants by Phoneme (/b/, /p/), Language Background (Native English, Native Mandarin-High, Native Mandarin-Low), and Context condition (No context, Context). The figure suggests that normalized durations were longer for the vowels preceding /b/ than those preceding /p/. However, this difference in the preceding vowel durations seems to be larger for Native English talkers' productions than for Native Mandarin-High talkers' productions; mean normalized vowel duration for /b/ – that for /p/ was 0.11 (Native English), 0.08 (Native Mandarin-High), 0.04 (Native Mandarin-Low). The linear mixed-effects regression model included Phoneme, Language Background group, as well as interactions between the two factors as fixed effects (see Supplementary Material for the model syntax and summary of the results). There was a significant effect of Phoneme (β = 0.08, t = 4.98, p < 0.001). The effect of Phoneme interacted with each Language Background group comparison: Phoneme × Native English vs. Native Mandarin-High (β = 0.07, t = 4.65, p < 0.001), Phoneme × Native Mandarin-High vs. Native Mandarin-Low (β = 0.07, t = 4.5, p < 0.001). This indicates that talkers produced longer vowels before /b/ than those before /p/. This difference in normalized vowel

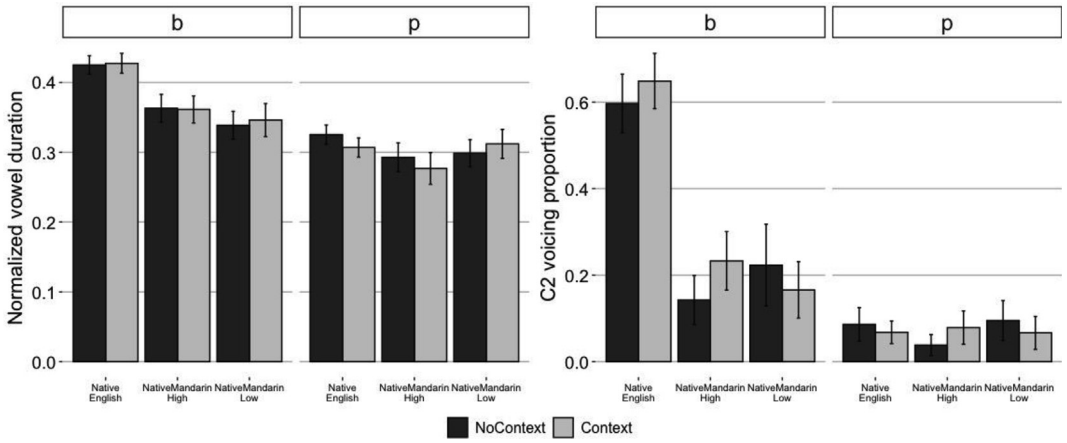

Fig. 2. Mean normalized durations of the vowels preceding the word-final target consonants (Left panel) and mean voicing proportions of the target consonants (Right panel), by Phoneme (/b/, /p/), Language Background group (Native English, Native Mandarin-High, Native Mandarin-Low), and Context condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

durations was larger for Native English talkers' productions than for Native Mandarin-Higher talkers' productions, and for Native Mandarin-Higher talkers' productions than for Native Mandarin-Low talkers' productions. A post-hoc Tukey test showed that the effect of Phoneme was significant for the Native English group (β = −0.11, SE = 0.02, t = −6.42, p < 0.0001), Native Mandarin-High group (β = −0.08, SE = 0.02, t = −4.36, p = 0.0001), and Native Mandarin-Low group (β = −0.04, SE = 0.02, t = −2.21, p = 0.033).

The left panel of Fig. 2 also suggests different patterns for normalized durations of vowels preceding /b/ and /p/. That is, although normalized durations of the vowels preceding /b/ differed for different Language Background groups, this Language Background-based difference was much smaller for normalized vowel durations preceding /p/. Further, there was a tendency for Native English and Native Mandarin-High talkers to shorten the vowel durations preceding /p/ in Context conditions compared to No Context conditions; whereas talkers generally did not differentiate vowel durations preceding /b/ in different conditions. The linear mixed-effects regression models, used to analyze normalized vowel durations as the dependent variable separately for /b/-targets and /p/-targets, included Context condition (No Context, Context), Language Background group, as well as interactions between the two as fixed effects (see Supplementary Materials for the model syntax and summary of the results). For the model with /b/-targets, there were significant effects of Language Background: Native English vs. Native Mandarin-High (β = 0.1, t = 7.48, p < 0.001), Native Mandarin-High vs. Native Mandarin-Low (β = 0.07, t = 4.72, p < 0.001). This indicates that normalized durations of vowels preceding /b/ were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The main effect of Context condition was not significant (β = 0.003, t = 0.4, p = 0.69). The effect of Context condition did not interact with the Native English vs. Native Mandarin-High comparison (β = −0.003, t = −0.15, p = 0.88), or with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = −0.01, t = −0.71, p = 0.48). This indicates that the effect of Context condition did not differ among different Language Background groups.

For the model with /p/-targets, there was a significant effect of the Native English vs. Native Mandarin-High comparison (β = 0.03, t = 2.04, p < 0.05). This indicates that normalized vowel durations were shorter for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Context condition was not significant (β = −0.007, t = −0.39, p = 0.7). However, there was a marginally significant interaction between Context condition and Native Mandarin-High vs. Native Mandarin-Low comparison (β = −0.03, t = −1.97, p = 0.05). This indicates that the effect of Condition was different between the two Language Background groups' productions; Native Mandarin-High talkers tended to decrease the normalized vowel durations in Context conditions compared to No Context conditions, though Native Mandarin-Low talkers did not. A post-hoc Tukey test showed that the effect of Context condition was not significant for any of the Language Background groups: Native English (β = 0.02, SE = 0.02, t = 0.86, p = 0.41), Native Mandarin-High (β = 0.02, SE = 0.02, t = 0.55, p = 0.59), Native Mandarin-Low (β = −0.008, SE = 0.02, t = −0.38, p = 0.71).

These results suggest that talkers distinguished word-final /b/ and /p/ with preceding vowel durations; normalized durations were longer for the vowels preceding /b/ than for those preceding /p/. This difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. This difference in normalized vowel durations for /b/- vs. /p/-targets among different talker groups was largely influenced by how talkers produced /b/-targets as compared to how they produced /p/-targets. That is, normalized vowel durations preceding /b/ were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. However, talkers did not manipulate normalized durations of the preceding vowels differently in different conditions (No Context, Context), either for /b/ or /p/. Though Native Mandarin-High and Native Mandarin-Low talkers manipulated the normalized vowel durations for /p/-targets in different directions (i.e., Native Mandarin-High talkers shortened the vowel durations in Context conditions compared to No Context conditions while Native Mandarin-Low

talkers slightly increased the vowel durations in Context conditions), these differences between Conditions were not statistically significant.

3.2.2. C2 voicing proportion

The right panel in Fig. 2 shows the mean C2 voicing proportions by Phoneme (/b/, /p/), Language Background (Native English, Native Mandarin-High, Native Mandarin-Low), and Context condition (No context, Context). The figure suggests that the C2 voicing proportions were larger for /b/ than for /p/ for Native English talkers' productions, but this difference was much smaller for Native Mandarin talkers' productions; mean C2 voicing proportion for /b/ – mean C2 voicing proportion for /p/ was 0.55 (Native English), 0.13 (Native Mandarin-High), 0.11 (Native Mandarin-Low). The linear mixed-effects regression model included Phoneme, Language Background group, as well as interactions between the two factors as fixed effects (see Supplementary Materials for the model syntax and summary of the results). The results showed a significant effect of Phoneme (β = 0.26, t = 11.56, p < 0.001). The effect of Phoneme interacted with each Language Background group comparison: Phoneme × Native English vs. Native Mandarin-High (β = 0.57, t = 9.59, p < 0.001), Phoneme × Native Mandarin-High vs. Native Mandarin-Low (β = 0.3, t = 4.48, p < 0.001). This indicates that the size of difference in C2 voicing proportions between /b/ and /p/ differed for different Language Background groups. The difference was larger for Native English talkers' productions than for Native Mandarin-Higher talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. A post-hoc Tukey test confirmed that the effect of Phoneme was significant for all the Language Background groups' productions: Native English (β = −0.55, SE = 0.03, t = −16.06, p < 0.0001), Native Mandarin-High (β = −0.13, SE = 0.04, t = −2.95, p = 0.004), Native Mandarin-Low (β = −0.11, SE = 0.04, t = −2.63, p = 0.01).

The right panel of Fig. 2 also suggests different patterns for C2 voicing proportions for each consonant (/b/ and /p/). For /b/, Native English talkers and Native Mandarin-High talkers increased the C2 voicing proportions from No Context to Context conditions; while Native Mandarin-Low talkers changed C2 voicing proportions in the opposite direction. However, the patterns for /p/ suggest that neither the Language Background nor Context condition influenced C2 voicing proportions.

The linear mixed-effects regression models, used to analyze C2 voicing proportions as the dependent variable separately for /b/-targets and /p/-targets, included Context condition (No Context, Context), Language Background group, as well as interactions between the two as fixed effects (see Supplementary Material for the model syntax and summary of the results). For the /b/-target model, there were significant effects of Language Background groups: Native English vs. Native Mandarin-High (β = 0.58, t = 9.26, p < 0.001), Native Mandarin-High vs. Native Mandarin-Low (β = 0.28, t = 4.05, p < 0.001). This indicates that C2 voicing proportions for /b/ were larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The effect of Context condition was not significant (β = 0.04, t = 1.76, p = 0.08); however, it interacted with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = 0.16, t = 2.33, p < 0.05). This indicates that the effect of Context condition was different between the Native Mandarin-High and Native Mandarin-Low groups' productions. The effect of Context condition did not interact with the Native English vs. Native Mandarin-High comparison (β = 0.06, t = 0.96, p = 0.34), indicating that the effect of Context condition did not differ between the Native English and Native Mandarin-High groups. A post-hoc Tukey test showed that the effect of Context condition was significant for the Native English group (β = −0.07, SE = 0.03, t = −2.03, p = 0.045), and the Native Mandarin-High group (β = −0.09, SE = 0.04, t = −2.09, p = 0.039), but not for the Native Mandarin-Low group (β = 0.04, SE = 0.04, t = 0.87, p = 0.39). For the /p/-target model, none of the Context condition or Language Background effects were significant (see Supplementary Material).

These results suggest that talkers distinguished word-final /b/ and /p/ with voicing proportions of the target consonants (C2); C2 voicing proportions were larger for /b/ than for /p/. This difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. This difference among different talker groups was influenced by how talkers produced /b/, not /p/. That is, voicing proportions of /b/ were larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. Furthermore, Native English talkers and Native Mandarin-High talkers manipulated C2 voicing proportions in different conditions (No Context vs. Context), and they did this differently for /b/ and /p/. That is, these talkers increased voicing proportions from No Context to Context conditions for /b/, but not for /p/. However, Native Mandarin-Low talkers did not manipulate C2 voicing proportions differently in different conditions.

4. Discussion

In the present study we examined acoustic enhancements produced by native and non-native talkers, when the communication context implicitly signaled that such enhancements were necessary. The present study demonstrates that talkers' target language experience impacts segmental enhancements of English sound contrasts in different ways depending on whether or not the sound contrast exists in non-native talkers' native language. That is, talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) did not impact how the talkers enhanced the English contrast that also exist in non-native talkers' native language (i.e., L1L2 consonant contrast) but it did for the English contrast that do not exist in non-native talkers' native language (i.e., L2-only consonant contrast). In the following sections, we discuss native and non-native talkers' segmental enhancement patterns in relation with the type of acoustic cues involved in the enhancements.

4.1. The effect of target language experience on contextually-relevant segmental enhancements

For the L1L2 consonant contrast (/p/-/b/ in word-initial position), non-native talkers of higher- and lower-proficiency enhanced the contrast by increasing the VOTs of /p/ as well as native English talkers did. These production patterns are in line with previous studies demonstrating that native English talkers exaggerate voiceless onset plosive VOTs when a voiced competitor is contextually present[5] (Baese-Berk & Goldrick, 2009; Buz et al., 2014, 2016; Kirov & Wilson, 2012). The current results further extend these findings to non-native talkers, demonstrating that non-native talkers of higher- and lower-proficiency use the same strategies to enhance the contextually-relevant voicing contrast in word-onset position, and they do so to a similar extent as native talkers do. These results suggest that manipulating acoustic properties to enhance a familiar non-native contrast is possible even for talkers with limited language proficiency (e.g., lower-proficiency talkers). Interestingly, native Mandarin speakers make larger baseline differences between the short- and long-lag stops than the native English speakers. We hypothesize that this could be attributed to the talker's L1 experience. While both English and Mandarin have two voicing categories word initially – short- and long-lag stops – the long-lag stops in Mandarin are produced with longer VOT than the same stop category in English (e.g., Chao & Chen, 2008; Rochet & Yanmei, 1991). Therefore, it is possible that Mandarin talkers are transferring some of their linguistic knowledge from L1 to L2 when producing these contrasts.

In contrast, the ability to enhance the L2-only consonant contrast (/p/-/b/ in word-final position) differed for higher- vs. lower-proficiency non-native talkers. Specifically, lower-proficiency talkers showed a tendency to decrease voicing proportions of /b/ in Context conditions as compared to No Context conditions; whereas native English and higher-proficiency talkers increased the voicing proportions of /b/ from No Context to Context conditions to enhance the coda /p/-/b/ contrast. These results demonstrated that among the non-native talkers, higher-proficiency talkers were better able to enhance the non-native contrast that does not exist in their native language compared to lower-proficiency talkers, possibly suggesting that as talkers' target language proficiency develops they are better able to use an acoustic enhancement strategy that they are not familiar with from their native language experience.

Here, it is important to point out that the current results suggest that the ability to distinguish a certain contrast (e.g., /p/ vs. /b/ word finally) may be partially independent from the ability to further enhance the contrast. In fact, although lower-proficiency talkers were able to differentiate between word-final /p/ and /b/, they did not enhance the contrast successfully. This potential gap between the ability to *produce* vs. *enhance* a contrast was also seen in the results, where higher-proficiency talkers were much less successful than native English talkers at *producing* the word-final /p/-/b/ contrast, though they were as successful as native English talkers at *enhancing* the same contrast. Thus, native vs. non-native language status differentially impacted talkers' ability to *produce* vs. *enhance* a non-native contrast. Furthermore, as higher-proficiency talkers enhanced the word-final /p/-/b/ contrast (the non-native contrast that does not exist in native language) much more successfully than lower-proficiency talkers did, this suggests that knowing how to *enhance* a non-native sound contrast (e.g., knowing which acoustic cue to manipulate in what way), in addition to knowing how to *produce* the contrast in general, may be a part of what characterizes language proficiency.

The results in the present study also suggest that native talkers and higher-proficiency talkers have acquired context-dependent knowledge of the voicing contrast in English, and the articulatory control to further enhance the position-specific details of the particular contrast. That is, these talkers do not treat the contrast the same way in word-initial and word-final positions. These results have implications for the relatively understudied area of inquiry regarding second language acquisition of allophonic variations. Particularly, previous work has shown that second language learners' use of allophonic variations is different from that of native talkers, and this may be impacted by language experience (e.g., Shea, 2014; Shea & Curtin, 2011; Vokic, 2010). It is also the case that previous work in this area has not directly addressed the issue of acquisition of novel phonotactic patterns in second languages. While there is a large body of literature on acquisition of novel phonotactic patterns (e.g., Warker & Dell, 2006), relatively little of this work is situated in the work on L2 acquisition (though see Steele et al., 2015 for a counter example). The current results contribute to these lines of studies by demonstrating that higher-proficiency talkers are able to enhance the voicing contrast in both word-initial and word-final positions (though lower-proficiency talkers were only able to enhance the contrast in word-initial position), highlighting the role of talkers' target language experience in their ability to manipulate important acoustic cues to enhance allophonic variations.

It should also be noted that 'proficient' talkers' (i.e., native English and higher-proficiency non-native talkers) use of acoustic cues to *enhance* the coda voicing contrast was different from those used to *produce* the contrast in general. That is, these talkers differentiated the coda /p/-/b/ contrast using both preceding vowel durations (i.e., longer vowels before /b/ than before /p/) and target consonant voicing proportions (i.e., larger voicing proportions for /b/ than for /p/); while they used preceding vowel durations to a much lesser extent than target consonant voicing proportions to further enhance the contrast, which is consistent with previous results (Goldrick et al., 2013). Though it is an open question why vowel durations preceding the target consonants are not utilized to enhance the coda voicing contrast, it is possible that talkers relied on the preceding vowel duration to different degrees for enhancing different segments involved in the coda voicing contrast. Specifically, higher-proficiency talkers and native talkers used the strategy of decreasing the vowel durations before /p/ (though the decrease in preceding vowel durations was not statistically significant), though for /b/, they used the strategy of increasing the consonant voicing proportions instead of manipulating vowel durations. This is in line with native English talkers' hyperarticulation patterns of a coda fricative voicing contrast (/s/-/z/:

[5] Note the lack of enhancement for voiced initial plosives is also consistent with previous literature (e.g., Goldrick, Vaughn, and Murphy, 2013; Nielsen, 2011) demonstrating that while talkers shift the voice onset time of voiceless plosives in English, they often do not shift the voice onset time of phonologically voiced plosives (i.e., "short-lag" or "zero VOT" stops).

Seyfarth et al., 2016), suggesting that the way coda voicing contrast is implemented could differ for voiced vs. voiceless end of the contrast across different manners of articulation. The current results further suggest that higher-proficiency non-native talkers could implement strategies for enhancing word-final voiced vs. voiceless sounds as native English talkers do, though these talkers may rely on one type of strategy more than the other in order to enhance the contrast.

4.2. Nature of contextually-relevant speech enhancements

The current results demonstrated the contextually-relevant enhancements observed here were quite targeted, highlighting the specificity of the acoustic modifications to enhance contrastive features of non-native sounds. In other words, while hyperarticulation is often associated with longer duration and expansion of vowel space (e.g., in clear speech: Bradlow et al., 2003; Picheny et al., 1986; Smiljanic & Bradlow, 2008), here we see that segmental contrast-enhancement can sometimes result in shortening of durations rather than lengthened duration. For example, in order to enhance the word-final /p/-/b/ contrast (L2-only consonant contrast), native English and higher-proficiency talkers tended to shorten the normalized durations of vowels preceding /p/s. Therefore, these results support the claim that types of contrastive hyperarticulation are not necessarily limited to elongation of segments, but can also involve shortening of durations or centralization of vowels in order to enhance specific contrasts (e.g., Leung et al., 2016; Seyfarth et al., 2016; Wedel, Nelson, & Sharp, 2018). Furthermore, the current results provide evidence that such targeted modifications can be found in productions of non-native talkers (higher-proficiency talkers) when potential ambiguity is signaled in the context.

Because the context-specific speech enhancements in the current study were examined using a paradigm that signals potential communicative difficulty in the context, one might wonder to what extent the enhancements were listener-driven. That is, native and non-native talkers' speech modifications implemented to enhance certain contrasts may have been driven not only by talkers' intention to be better understood by listeners (listener-oriented) but also by talkers' internal processing of the target lexical items (talker-oriented). Previous studies have suggested several theoretical accounts for contrastive hyperarticulation. One explanation is that contrastive hyperarticulation is based on talkers' modeling of listeners' communicative needs (perceptual monitoring, or communication-based accounts: Baese-Berk & Goldrick, 2009; Buz et al., 2016). That is, talkers modify phonetic characteristics of their productions based on their understanding of what their listeners understand or know in the communication. However, another explanation is that contrastive hyperarticulation is facilitated by lexical competition during production planning (production-internal account; see Baese-Berk & Goldrick, 2009 for detailed discussion). That is, the presence of phonologically similar words in the same context as the target words increases the difficulty of phonological encoding during planning, and this causes higher activation of the target words, resulting in hyperarticulation. This theoretical account suggests that contrastive hyperarticulation is talker-driven, originating from lexical and phonological planning processes of the talker.

The current study was not designed to differentiate these types of explanations for contrastive hyperarticulation, and we find that the current results could be compatible with both of these explanations. Particularly, it is possible that talkers hyperarticulated the contrasts because target words were presented with their minimal-pair neighbors in the same context, increasing lexical competition between those words. However, because some of the enhancements made by native and non-native talkers were quite targeted to enhance specific contrasts (e.g., lengthening of segment durations that is independent of overall word durations or shortening of segment durations), it is also plausible that speech enhancements observed in the current study were at least to some extent driven by talkers' intention of increasing perceptual distance of the contrasts for the listener. Thus, based on these results, we suggest that talker-oriented and listener-oriented explanations for contrastive hyperarticulation may not be exclusive of one another, and they could work in concert to characterize talkers' production patterns. In fact, some previous results suggest that production-internal processing (e.g., lexical neighborhood-density effects) and listener-oriented processing (e.g., effects of clear speech instructions) can both impact talkers' hyperarticulation, and these effects may be independent of one another (e.g., Scarborough, 2010; Scarborough & Zellou, 2013). Thus, native and non-native talkers' contextually-relevant speech enhancements observed in the current study could be explained by combinations of talker-driven and listener-driven processes. However, some aspects of these explanations may not directly apply to non-native talkers' contrastive speech enhancements. For example, previous studies have suggested that native talkers' contrastive hyperarticulation for words with minimal pairs (e.g., *cod vs. god*; as compared to those without minimal pairs: *cop vs. *gop*) can occur without the overt presence of minimal-pair neighbor in the same context as the target word (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Wedel et al., 2018). However, such hyperarticulation driven by production-internal lexical processing may not necessarily occur for non-native talkers' productions when the talkers do not realize that some non-native words have minimal pairs and some do not, or that some non-native words have higher-neighborhood density than others. Furthermore, whether non-native talkers are able to implement the intended enhancements via their articulatory control (e.g., increasing voicing proportions of word-final stop or fricative consonants: Seyfarth et al., 2016, current results) would be a separate question from how their hyperarticulation is induced (by lexical processing internal to talkers' production system and/or by talkers' modeling of listeners' communicative needs). Thus, it is an open question whether mechanisms underlying contextually-relevant contrastive enhancements are similar for talkers of different linguistic backgrounds.

5. Conclusions

This study examined acoustic characteristics of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency. When the potential communica-

tion difficulty was signaled in the communication context, talkers made acoustic adjustments to enhance sound contrasts. Characteristics of these enhancements were also affected by talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) as well as by the type of the enhancements. Particularly, though we did not observe the effect of talkers' target language experience on the size of acoustic modifications for the non-native contrast that both native and non-native talkers are familiar with (i.e., non-native contrast that exists in non-native talkers' native language: L1L2 consonant contrast), we found the effect of target language experience for the non-native contrast that does not exist in non-native talkers' native language (L2-only consonant contrast). The current findings add to the growing body of work showing that talkers are able to accommodate phonetic characteristics of their productions based on the potential communication difficulty signaled in the context, and that these findings can be extended to the productions of higher- and lower-proficiency non-native talkers. Furthermore, non-native talkers' ability to enhance a non-native contrast improves as their target language proficiency level develops. Knowing how to enhance a specific sound contrast, in addition to knowing how to produce the contrast in general, may be a part of what characterizes language proficiency, though such effect of proficiency level could differ depending on the type of acoustic manipulations required to enhance the contrast.

CRediT authorship contribution statement

Misaki Kato: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Melissa M. Baese-Berk:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This work was funded by National Science Foundation Grants BCS-194173 to MMBB and MK, BCS-1734166 to MMBB, 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 MK. We would like to thank Ilsa Trummer, Robin Rogers, and Paolo Daniele at the University of Oregon for their assistance with subject recruitment and Sarah Steindorf and Brandon Lasala for their help with acoustic analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2021.101099.

References

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*(4), 527–554.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Iot International, 5*(9/10), 341–345.

Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII (phonology and phonetics)* (pp. 241–273). Berlin: Mouton de Gruyter.

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

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for learning-impaired children: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research, 46*, 80–97.

Brière, E. J. (1966). An investigation on phonological interference. *Language, 42*, 768–796.

Buz, E., Jaeger, T. F., & Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyperarticulation. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 36, No. 36).

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language, 89*, 68–86.

Chao, K. Y., & Chen, L. M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics & Chinese Language Processing, 13*(2), 215–232.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America, 134*(2), EL172–EL177.

Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic–phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. *Journal of Phonetics, 40*(3), 509–520.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America, 130*(4), 2139–2152.

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language, 81*, 72–90.

Kirov, C., & Wilson, C. (2012). The specificity of online variation in speech production. In Proceedings of the 34th Annual Conference of the Cognitive Science Society (pp. 587–592). Austin, TX.

Kuznetsova, A., Brockoff, P. B., & Christensen, R. H. B. (2016). Tests in Linear Mixed Effects Models. R package lmerTest version 2.0-30. Comprehensive R Archive Network (CRAN).

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Leung, K. K. W., Jongman, A., Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *Journal of the Acoustical Society of America, 140*(1), 45–58.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electrical hearing. *Journal of the Acoustical Society of America, 116*, 2374–2383.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America, 125*(6), 3962–3973.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142.

Ohala, J. (1994). Acoustic study of clear speech: A test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody* (pp. 75–89).

Oviatt, S., Levow, G.-A., Moreton, E., & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human-computer error resolution. *Journal of the Acoustical Society of America, 104*(5), 3080–3098.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 28*(1), 96–103.

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.

Rochet, B. L., & Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustics, 19*(4), 105–106.

Rogers, C. L., DeMasi, T. M., & Krause, J. C. (2010). Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustical Society of America, 128*(1), 410–423.

Scarborough, R. (2010). Lexical and conceptual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.). *Papers in Laboratory Phonology* (Vol. 10, pp. 557–586). Berlin: de Gruyter.

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. Journal of the Acoustical Society of America, 121(5), 3044.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of the Acoustical Society of America, 134*(5), 3793–3807.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics, 41*(3-4), 249–263.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime: User's guide. Psychology Software Incorporated.

Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal-American Academy of Audiology, 7*, 212–218.

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America, 139*(2), EL31–EL37.

Shea, C. E. (2014). Second language learners and the variable speech signal. *Frontiers in Psychology, 5*, 1–3.

Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research, 27*(2), 229–250.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91–113.
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.
Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *Journal of the Acoustical Society of America, 130*(6), 4020–4031.
Steele, A., Denby, T., Chan, C., & Goldrick, M. (2015). *Learning non-native phonotactic constraints over the web.* Glasgow, United Kingdom: The University of Glasgow.
Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*(3), 163–178.
Tuomainen, O. T., & Hazan, V. (2018). Investigating clear speech adaptations in spontaneous speech produced in communicative settings. In M. Gósy & T. E. Gráczi (Eds.), *Challenges in analysis and processing of spontaneous speech* (pp. 9–25). Research Institute for Linguistics for the Hungarian Academy of Sciences.
Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 207–235). Malden, MA: Blackwell Publishers.
Vokic, G. (2008). The role of structural position in L2 phonological acquisition: Evidence from English learners of Spanish as L2. *Foreign Language Annals, 41*(2), 347–363.
Vokic, G. (2010). L1 allophones in L2 speech production: The case of English learners of Spanish. *Hispania*, 430–452.
Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 387.
Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language, 100*, 61–88.