What accounts for foreign accent reduction in singing?

Kit Ying Chan[1], Michael D. Hall[2], and Riley Herr[2]

[1]Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong

[2]Department of Psychology, James Madison University, Harrisonburg, USA

Author Note

Correspondence concerning this article should be addressed to Michael Hall, Department of Psychology, MSC 7704, James Madison University, Harrisonburg, VA 22807, USA. Email: hallmd@jmu.edu

Abstract

Second-language learners often sound less foreign-accented when they sing than speak. This study explored whether production rate, particularly vowel lengthening, represents the key factor for accent reduction in song. Stimuli were lyrical lines extracted across spoken and sung versions of popular songs produced by the most accented male and female non-native speakers from the NUS-48E corpus (Duan, et al., 2013). Lengthened versions also were included, either throughout the signal, or for vowels alone, by slowing productions to .67 of the original rate while maintaining fundamental frequency (F_0). A total of 33 male and 37 female native English speakers were randomly assigned to rate the level of accentedness for excerpts produced by either male or female speakers..

Sung excerpts were perceived as less accented than speech. Unexpectedly, both sung and spoken original excerpts were less accented than both lengthened versions, which provided similar ratings. This suggests that vowel lengthening cannot consistently be responsible for accent reduction in songs. Acoustic analyses revealed a higher median F_0 in songs. Median F_0 from read and sung excerpts negatively correlated with accentedness ratings, regardless of the speaker's self-identified gender. This suggests that the spreading of adjacent harmonics as F_0 increases may under-represent the spectrum, making it difficult for listeners to assess whether vocal tract resonances align with those typically generated by native-language producers.

Keywords: accent reduction; songs; rate; vowel duration; fundamental frequency; foreign accent

What accounts for foreign accent reduction in singing?

Many music artists from the United Kingdom, Australia, and New Zealand sing with an American accent, but they speak in their regional accent (Gibson, 2010; Morrissey, 2008; Simpson, 1999; Trudgill, 1983). Regardless of whether pop singers imitate the accent deliberately (Trudgill, 1983) or not (Beal, 2009; Gibson, 2010), they are professional singers with extensive vocal-motor training and enhanced vocal-motor control. There is evidence that good singing ability predicts good speech imitation performance in a familiar second language (L2) and a completely unintelligible foreign language (Christiner & Reiterer, 2013), as well as the ability to fake a foreign accent in the native language (Coumel, Christiner, & Reiterer, 2019). Trained singers' greater vocal-motor flexibility and control enable them to store more accurate representation of phonemes (Coumel et al., 2019), produce better sounds, and cover a larger articulation space (Reiterer, Hu, Sumathi, & Singh, 2013).

Accent neutralization in songs is not only found in professional singers regarding regional accent, but also in untrained and inexperienced singers regarding foreign accent. Foreign accentedness has been found to be reduced in L2 learners without extensive vocal training when they sing compared to when they speak (Hagen, Kerkhoff, & Gussenhoven, 2011; Mageau, Mekik, Sokalski, & Toivonen, 2019). These findings suggest an inherent effect in singing that might mask certain markers of accent. Foreign accent in speech is characterized by distinct changes in rhythm (Polyanskaya, Ordin, & Busa, 2016; Riazantseva, 2001), intonation (Munro & Derwing, 1995, 2001; Polyanskaya et al., 2016; Sereno, Lammers, & Jongman, 2016; van Maastricht, Zee, Krahmer, & Swerts, 2021), and vowel quality and duration (Chan, Hall, & Assgari, 2017; Munro, 1993; Wade, Jongman, & Sereno, 2007). It is possible that the melody and the rhythm of a song impact and constrain duration and F_0 patterns to the extent that the

usual patterns and variation that are present in speech are not apparent (Hagen et al., 2011). For example, singers, regardless of whether they are novices or have received professional training, tend to open the jaw more as pitch increases, resulting in more open vowels for higher notes (Austin, 2007; Sundberg & Skoog, 1997). Moreover, vowels are often elongated in songs. Singers might be forced to use the easiest vowel shape to sustain the vowels, resulting in changes in vowel quality. Hence, it is likely that the identifying features of accents can be masked by singing.

Rather than focusing on vocal control to fit a song's rhythm and melody, the current study investigated the acoustic parameters that are linked to accent reduction in song. A previous study examined the role of syllable durations and F_0 patterns in signaling foreign accents in speech and in songs (Hagen et al., 2011). Spoken and sung versions of sentence-long lyrics of familiar songs by Dutch speakers of English as L2, as well as by native speakers of English, were recorded. The spoken versions from Dutch speakers were then manipulated by 1) matching segmental durations of non-native productions to the corresponding productions by one native speaker, 2) making F_0 monotonic in non-native productions, and 3) matching segmental durations and making F_0 monotonic. The stimuli were rated by native English listeners on native accent authenticity. Productions by Dutch speakers were rated to have higher authenticity when the same lyrics were sung rather than when they were spoken; no such significant difference was found in the spoken and sung productions by the native speakers. Also, the monotonic version with English durations was judged to have higher authenticity compared to the monotonic version with original durations and the version with English durations and original F_0. It was concluded that the rhythm and melody of songs impose certain segmental durations and F_0 variation on sung text, leaving the two intonational markers unavailable to signal foreign

accentedness (Hagen et al., 2011). However, this study only manipulated segmental duration and F_0 in the spoken stimuli to indirectly infer their role in signaling foreign accentedness in songs. A direct acoustic comparison on these features for spoken and sung versions by the native speakers and Dutch speakers is required to confirm that singing neutralizes these features.

The influence of singing on the durations and F_0 (and thus, presumably, pitch) for vowels also was examined through acoustic analyses of speech and songs from a group of native and non-native English speakers without musical training (Mageau et al., 2019). For non-native speakers, songs again were rated as less foreign-accented than speech. Furthermore, for both native and non-native speakers, vowel duration was longer, and F_0 was higher, within songs relative to speech. In English, vowel duration tends to be longer preceding voiced consonants (e.g., /ʌ/ in *buzz)* than voiceless consonants (e.g., /ʌ/ in *bus),* and it is important for cueing the voicing of word-final consonants (Kluender, Diehl, & Wright, 1988; Peterson & Lehiste, 1960). The distinction in vowel length before voiced and voiceless consonants was greater in speech than in songs for non-native speakers. This voiced-voiceless distinction in vowel duration between native and non-native speakers also differed in speech, but not in song.

Taken together, findings from Hagen et al. (2011) and Megeau et al. (2019) suggest that foreign accent might be reduced by vowel durations within songs. Although vowel duration does not change word meaning in English, it is an important cue for vowel height (Toivonen et al., 2015) and consonant voicing (Kluender et al., 1988; Peterson & Lehiste, 1960). It also has been shown to be a crucial foreign-accent marker in speech (Munro, 1993; Tajima, Port, & Dalby, 1997). A larger difference in vowel duration between English minimal pairs ending in /t/ and /d/ by non-native speakers was judged to be less foreign-accented and showed a larger voicing effect (Flege, 1993). Vowels are typically lengthened in songs to carry the pitch in each syllable to

comply to the rhythmic nature of song (Mageau et al., 2019). It therefore is possible that lengthening vowel duration could be the key factor contributing to the reduction of perceived foreign accents in songs.

The current study aimed to test this hypothesis by lengthening vowel durations in corresponding spoken and sung lyrics produced by a group of non-native speakers of English and then measure the impact of that lengthening on perceptual ratings of foreign-accentedness. Speech and song samples were presented at their original production rates, as well as at a distinctly slower rate for all vowels. A third condition that slowed both vowels and consonants also was included as a control. Consistent with the aforementioned literature, it was hypothesized that sung samples by a non-native speaker would be rated as less foreign-accented compared to when the same text was read. It was additionally hypothesized that vowel lengthening in both non-native speech and song might reduce perceived foreign accentedness compared to the corresponding original productions. An acoustic analysis of F_0 values from the original read and sung excerpts also was conducted to examine the potential role of F_0 in accent reduction.

Method

Participants

Ninety-four native speakers of English (45 female and 49 male) were recruited from Amazon Mechanical Turk to participate in this study with a small amount of compensation. Sixteen male and 8 female respondents were excluded, as they reported English not as their first language, not having normal vision or corrected-to-normal vision, or (ever) having a hearing or speech disorder. After screening, seventy participants (37 female and 33 male) were retained. About 63% of these participants reported to be born in the United States, followed by 25.7% from India, 4.3% from the United Kingdom, 2.9% from Italy, 2.9% from France, and 1.4% from

Canada. The mean ages were 33.7 years old (SD = 9.5; ranging from 22-63 years) for the males and 32.4 years old (SD = 9.9; ranging from 19 – 63 years) for the females. Participants reported no history of hearing or speech disorder. About 43% of the participants spoke another language apart from English. About 51% of the participants had visited or lived in foreign counties that speak a language other than English for more than 3 months. Half of the participants reported having family members or close friends with a foreign accent. Seventy percent of the participants reported having regular contact with non-native speakers of English.

Materials

Participant experiences with languages were determined using the *Participant Language Background Questionnaire*, which is provided in Appendix A. This questionnaire includes questions about the city and place of birth, sex, age, native language, history of hearing or speech disorder, other language ability, and contact with non-native speakers of English. Individual responses from the questionnaire also were used to screen qualifying participants.

Stimuli

Audio recordings (with a sampling rate of 44.1 kHz and 16-bit depth resolution) of the first two lines of sung and spoken lyrics of English popular songs were chosen from the NUS-48E Sung and Spoken Corpus, which was developed by the Sound and Music Computing Laboratory at National University of Singapore (Duan, Fang, Li, Sim, & Wang, 2013). Excerpts were of sung and spoken lyrics from 4 English popular songs produced by the 6 most accented (of 12) non-native speakers/singers. The excerpts reflected 3 men (#7, 8, and 12 from the corpus), two of whom have a mild Singaporean accent (#7 and 12) plus one who has a northern Chinese accent (#8), and 3 women (#4, 5, and 6 from the corpus), two of whom have a mild Malay accent (#4 and 6) plus one who has a pronounced Malay accent (#5). Accents were

determined by the original authors. Song selections were generally matched across 1 male speaker/singer and a corresponding female speaker/singer (#12 with #4, and #8 with #5). The one exception was for speakers/singers #7 and #6, who shared two of the four selections. There were 2 delivery modes (sung/read) x 6 singers (3 men, 3 women) x 4 songs x 3 production rates (original, slow, and vowel-slow), resulting in 144 stimuli. For all original stimuli, there was no significant difference between male (M = 4.78, SD = 2.17) and female singers/speakers (M = 4.66, SD = 2.21) with respect to overall *stimulus duration*, F (1, 44) = .066, p = .799. For all original stimuli, there was also no significant difference between male (M_{total} = 2.51, SD_{total} = 1.54; $M_{average}$ = 241, $SD_{average}$ = 188) and female singers/speakers (M_{total} = 2.57, SD_{total} = 1.65; $M_{average}$ = 242, $SD_{average}$ = 164) with respect to *total vowel duration per production* (i.e., the summed duration of all vowels within the production), F (1, 44) = .046, p = .831, or *mean vowel duration per production* (i.e., the total vowel duration divided by the number of vowels within the production), F (1, 44) < .001, p = .990.

All stimulus editing was performed using Adobe *Audition* 3.0 (Adobe Systems Inc., 2007). In addition to the "original", unaltered productions, alternative sets of productions at a slower rate (.67 times the original rate while maintaining F_0) were generated using the software's time-stretch routine. One set of "slow" productions reduced the rate for the entire signal. Another set of "vowel-slow" variants instead reduced the production rate for only the vowels. Vowel positions within these variant productions were determined to the nearest millisecond according to phonetic transcriptions provided by authors of the corpus. In a few instances, secondary adjustments were necessary at zero-crossings to ensure click-free playback. All stimuli were equated for average RMS amplitude in an attempt to minimize overall differences in loudness while still maintaining important spectral relationships within any given recording.

Procedure

Participants from Amazon Mechanical Turk were presented a consent form electronically through *QuestionPro* (QuestionPro, 2021). They filled out a Participant Language Background Questionnaire, followed by an accentedness rating task. Each participant heard the 72 stimuli consisting of lyrics sung or spoken by either the three male or the three female non-native speakers/singers across the three different production rates. On each trial, upon hearing the entire stimulus, participants were instructed to rate their perceived level of foreign accentedness for the production using a 9-point scale (*1* = native-like - *9* = extremely strong foreign-accented). Participants then clicked on 'Next' on the screen to reveal the next trial. Participants were not given specific instructions as to whether they should listen to the stimuli using headphones or over loudspeaker(s). The presentation order of the 72 stimuli, consisting of excerpts from 4 songs with 2 delivery modes (sung/read) by 3 speaker/singers from the same self-identified gender at 3 production rates (original, slow, and vowel-slow), was completely randomized. Each stimulus was presented once only. The entire experiment, that is, consent, the background questionnaire, followed by the rating task, and debriefing, required about 20 minutes to complete.

Results

For a participant-wise analysis, accentedness ratings from each participant were pooled across stimulus items of the same delivery mode (read and sung) and production rate (original, slow, and vowel-slow). The resulting 6 mean accentedness ratings per participant were submitted to a 2 x 3 repeated measures analyses of variance (ANOVA) with delivery mode and production rate as factors. As the four songs were different for each speaker/singer, song was not analyzed as a factor. Speakers'/singers' self-identified gender was not analyzed as a factor at this stage. We only included speakers'/singers' gender as a factor in the item-wise F_0 analyses later when

the expected gender difference in F_0 values was relevant. All post-hoc, pair-wise comparisons of means reflected Bonferroni adjustments.

Figure 1 displays the mean ratings of foreign accentedness (along with corresponding standard errors) for read and sung excerpts as a function of production rate condition. Looking at Figure 1, it can be seen that sung excerpts (M = 4.53, SE = .163) were less accented than corresponding read excerpts (M = 5.73, SE = .149). This tendency was confirmed by a significant main effect of delivery mode, $F(1, 138) = 77.8$, $p < .001$, $\eta_p^2 = .53$.

Figure 1 also reveals a tendency for the original excerpts to be perceived as less accented than the two lengthened versions. This was confirmed by a significant main effect of production rate, $F(2, 138) = 7.98$, $p = .001$, $\eta_p^2 = .104$. Subsequent pair-wise comparisons further revealed that this main effect was due to perceived accentedness being significantly reduced for the original excerpts (M = 4.99, SE = .146); no significant differences were observed between mean ratings for the two lengthened conditions (slow: M = 5.21, SE = .142; vowel-slow only: M = 5.20, SE = .146).

To examine if vowel duration is the primary factor in accent reduction in songs, different durational measures from each original stimulus items were derived and subjected to item-wise analyses, including overall *stimulus duration*, *total vowel duration per production*, *mean vowel duration per production* , *non-vowel duration*, *proportion of vowels in the original stimulus* (i.e., total vowel duration divided by overall stimulus duration), and *durational ratio of vowels to consonants* (i.e., total vowel duration per production divided by non-vowel duration). These durational measures were submitted to a 2 x 2 analyses of variance (ANOVA) with delivery modes (sung/read) and speakers'/singers' self-identified gender (male/ female) as the between-items factors. Table 1 showed the means and standard deviation for the durational measures for

read and sung stimuli for each self-identified gender separately, as well as the *F* statistics for the significant difference between durations (for each given measure) for read and sung stimuli. All other main effects or interactions failed to reach statistical significance.

Correlations between several durational measures and accentedness ratings were conducted. A similar pattern of results was obtained with respect to correlations of ratings with overall stimulus duration, the summed duration of all vowels per production, and the mean duration of vowels within excerpts. As the mean duration of vowels within excerpts controlled for the fact that there were different numbers of vowels across the excerpts, we only report findings based upon that measure here. There was an opposing tendency for the perceived accentedness of the original excerpts alone to decrease with the time spent in vowel production. This is reflected in Figure 2, which provides a scatterplot of mean ratings of accentedness against the mean duration of vowels within excerpts in each original spoken and sung production as a function of the producer's self-identified gender. This pattern contributed to a significant negative correlation between accentedness rating and mean duration of vowels within excerpts, r (48) = -.431, p = .002. Corresponding correlations also were obtained for each self-identified gender, for males [r (24) = -.522, p = .009] and for females [r (24) = -.414, p = .044]. As can be seen in Figure 2, this correlation was due to the fact that vowel productions were necessarily longer for songs (as indicated by the unfilled symbols) relative to speech (filled symbols). No further relationships between mean vowel duration and accentedness ratings were obtained when analyses were further restricted to particular combinations of speech or song with a specific self-identified gender (i.e., for male speakers, r (12) = -.227, p = .478; for male singers, r (12) = .104, p = .748; for female speakers, r (12) = .320, p = .311; for female singers, r (12) = -.052, p = .872).

To investigate if differences in the rhythm of the production, in terms of the durational ratio of vowels to consonants within the production, are related to perception of accent as shown in non-native speech within previous studies (Polyanskaya & Ordin, 2019; Polyanskaya et al., 2016), a correlation was run. There was a significant negative correlation between the durational ratio of vowels to consonants and mean accentedness ratings for read and sung original stimuli, $r(48) = -.331$, $p = .021$. Figure 3 shows a scatterplot of mean ratings of accentedness against durational ratio of vowels to consonants within excerpts in each original spoken and sung production as a function of the producer's self-identified gender. As the durational ratio of vowels to consonants within the production increased, mean accentedness ratings decreased. This confirmed that the rhythm (i.e., the patterns of durational variability in speech intervals) is related to perceived accentedness not only in speech, but also in songs.

To examine if F_0 might represent an alternative, effective predictor of accent reduction in songs, F_0 values from the original stimuli, including both read and sung excerpts by both male and female speakers/singers, were evaluated through acoustic analyses using *Praat* sound analysis and synthesis software (version 6.1.51; Boersma & Weenink, 2022). For each stimulus file, the median fundamental frequency (F_0, in Hz) was determined for item-wise analyses. There was a total of 48 data points from excerpts of 4 songs with 2 delivery modes (sung/read) by 6 speaker/singers (3 males and 3 females). For an item-wise analysis, median F_0 values were submitted to 2 x 2 mixed effects ANOVAs with delivery mode (read and sung) as the within-items factor and speaker's/singer's gender (male and female) as the between-items factor. As the four songs were different for each speaker/singer, song was not analyzed as a factor.

Figure 4 displays a scatterplot between median F_0 values (Hz) obtained for speech and songs by both male and female speakers/singers at the original production rate. From Figure 4, it

can be seen that the median F_0 values for songs (M = 245, SE = 4.17) tended to be higher than those for speech (M = 189, SE = 3.05). This tendency was confirmed by a significant main effect of delivery mode, $F(1, 22) = 142$, $p < .001$, $\eta_p^2 = .866$. Figure 4 also reflects that female speakers/singers (M = 278, SE = 3.96) tended to have higher median F_0 values than male speakers/singers (M = 156, SE = 3.96), $F(1, 22) = 475$, $p < .001$, $\eta_p^2 = .956$. As a result, median F_0 measures from speakers/singers of each self-identified gender were separately subjected to correlation analyses with mean accentedness ratings.

Mean accentedness ratings were obtained for each of the 48 items (4 songs x 2 delivery modes x 6 speakers/singers) by pooling the ratings for the same items across participants. A scatterplot of mean ratings of accentedness against median F_0 values in each original spoken and sung production as a function of the producer's self-identified gender is depicted in Figure 5. As can be seen in the figure, there was a tendency for accentedness ratings to decrease as median F_0 increased. A significant negative correlation between mean accentedness ratings and median F_0 was obtained for read and sung stimuli from both male [$r(24) = -.638$, $p = .001$] and female speakers/singers [$r(24) = -.678$, $p < .001$]. When correlation analyses between mean accentedness ratings and median F_0 were conducted separately for read and sung stimuli for each self-identified gender, only a significant correlation was obtained for read stimuli from female speakers, $r(12) = -.693$, $p = .012$ (see filled circles in Figure 5). Correlations between mean accentedness ratings and median F_0 were not significant for sung stimuli from female speakers [$r(12) = -.171$, $p = .596$], read stimuli from male speakers [$r(12) = .238$, $p = .457$], and sung stimuli from male speakers [$r(12) = -.264$, $p = .407$].

Discussion

This study examined if vowel lengthening is a key factor contributing to accent reduction in songs by directly measuring the effect of lengthening on perceived accentedness in the songs and speech of non-native speakers of English. Consistent with previous findings in Hagen et al. (2001) and Mageau et al. (2019), the voices of the non-native speakers were rated as less accented when they sung relative to when they spoke. In contrast to our hypotheses, the original excerpts, both for speech and song, were rated as less accented than any of the lengthened excerpts.

There are several potential explanations for such an unexpected finding. For example, one might suggest that the increase in accentedness that occurred with lengthening reflected stimuli becoming more artificial, and thus, sounding less human. After all, a large amount of lengthening requires the duplication of many sample points, which, in large doses, can produce more robotic speech. However, this account seems quite unlikely given that our laboratory personnel deemed the stimuli as a set to sound quite natural. In fact, extensive anecdotal listening by all authors was undertaken to arrive at a determination of the selected value for the rate manipulations. This value was selected to provide a strong manipulation on the provision that the results still sounded consistently human and natural.

An alternative argument could be made that the observed effects could reflect, at least to a degree, differences in the rhythmic characteristics of the productions. Previous studies have revealed that deviations from native speech rhythm, in terms of durational ratios (e.g., between vocalic and other segments of speech), can increase the perceived accentedness of non-native speech (Polyanskaya & Ordin, 2019; Polyanskaya et al., 2016). Some evidence also suggests that

this impact of rhythm on accentedness in speech is greater than corresponding effects of rate manipulations alone (Polyanskaya et al., 2016).

Within the vowel-slow condition of the current investigation, lengthening the vowel durations by a flat rate of .67 of the original rate by definition disproportionately increased the lengths of vocalic portions of the utterances relative to the remaining segments in those signals, and thus, should have impacted durational ratios (in a manner that deviated from native norms). However, similar increases in accentedness were observed not only in the vowel-slow condition, but also within the slow condition. This occurred even though rate in the slow condition was adjusted equivalently across all segments of a given utterance, thereby maintaining the original durational ratios. Thus, it appears that, while rhythmic changes away from native norms represents an important contributing factor to the perception of foreign accent in speech, the major findings from the current investigation can be considered rate-specific effects. It is noteworthy that we were initially concerned during stimulus construction about potential rhythmic changes that might result in the vowel-slow condition, particularly for songs, where the result could conceivably compromise segmental durations relative to the original musical composition. Yet, in our estimation the synthesized results did not sound even slightly unnatural in their rhythms. Furthermore, corresponding increases in accentedness were obtained across both vowel-slow and slow conditions, suggesting that these effects were carried specifically by vowels.

Two other accounts could be easily argued to provide more viable alternative explanations. One possibility is that vowel lengthening is not the primary parameter responsible for foreign accent reduction in songs, as had been generally assumed. While the current investigation clearly appears to have obtained an effect of vowel lengthening, that effect was in

the opposite direction of what was expected. The average change in perceived accentedness also was quite small (about a quarter of a point in terms of the rating scale) and produced a modest/moderate effect size. Furthermore, while increasing average vowel length within the originally recorded stimuli actually predicted reduced accent (see Figure 2), a correlation with accentedness that was at least as strong, if not stronger, was obtained from median F_0 values (see Figure 5). Thus, it is certainly possible, at the very least, that any benefits of vowel durations in those stimuli to accent reduction might have been in conjunction with another parameter that represents at least an equally strong predictor of performance in that (original) condition.

A more conservative alternative is that there could be an optimal range of speaking/singing rates for native listeners, and thus, an optimal range of segmental duration values, that non-native speakers tend to exceed. According to this alternative, foreign accentedness would be perceived when speech or singing is produced that is either very fast or very slow relative to native norms. In the current investigation, the obtained negative correlation between accentedness and average vowel durations in the original stimuli is consistent with the notion that the more rapid vowel productions (or, minimally, productions with relatively few vowels) were perceived as more accented. Also, the non-native talkers/singers generally were perceived to have a strong foreign accent, and were producing samples of English that were already slow. As a result, further slowing through our stimulus manipulation generated very slow examples, which would make them likely to be perceived as even more accented. It is notable that this should not typically occur within recorded music production when there is instrumental accompaniment to singing, as the singer would be obliged to match the song's intended tempo and rhythmic structure, as conveyed by the instruments. An exception would be where instrumentalists naturally adjust their timing to match the singer when playing or recording as a

live band. However, no such opportunities occurred within the singing database that contributed to stimuli from the current investigation, as all singers produced acapella samples (i.e., in the absence of instrumental accompaniment).

It is acknowledged that the current investigation was not designed to differentiate between these two viable explanations of increased accentedness with slowed speech and song. Since the study was specifically targeting the effects of vowel lengthening, our rate manipulation occurred in only one (slowed) direction. Thus, further investigation is warranted to determine whether the major findings reported here indicate a reliance on a constrained set of rates in determining the strength of a perceived foreign accent and/or whether this occurs in conjunction with other, potentially more critical, stimulus parameters, such as F_0. For now, we can still conclude based upon the major findings from the current investigation that vowel lengthening does not automatically, or necessarily, reduce the perception of foreign accent across all natural speaking or singing rates.

A closer look at the relevant literature(s) reveals that this finding should not necessarily be considered terribly surprising. Although we believe that the current investigation represents the first to reveal an effect of slower production rates on increasing foreign accentedness for song, corresponding effects have been previously obtained for speech (e.g., see Munro & Derwing, 1998, 2001). It also has been well established that non-native speech is generally slower than native speech (Guion, Flege, Liu, & Yeni-Komshian, 2000; Lennon, 1990; Munro & Derwing, 1998). Speech from Mandarin speakers of English as L2 is rated as more accented when it is produced at slower-than-normal rates relative to natural rates for the given talkers (Munro and Drewing,1998). Furthermore, consistent with the range-based alternative offered above, accentedness was found to be reduced in non-native speech that was slightly faster than

typical rates produced by L2 learners, but slower than the corresponding rates by native speakers; in contrast, accentedness increased when given both very fast, and very slow, non-native speech (see Munro & Drewing, 2001). These collective findings led the authors to speculate that phonological errors might become more salient in non-native speech at slower rates, thereby increasing perceived foreign-accentedness. Similarly, our manipulations of slowing down the vowels only and the whole excerpts in this study might highlight the acoustic-phonetic deviations in the segments, increasing the perceived foreign-accentedness of both speech and songs.

Another factor that likely contributes heavily to the reduction of foreign accent in songs, and, at the very least, likely impacted the distribution of accentedness ratings for the original stimuli within the current investigation, is F_0. Acoustic analyses of the original songs and speech showed that songs had higher median F_0 values than speech for both male and female readers/singers, and there was a significant negative correlation between mean accentedness ratings and median F_0 values. This suggests that heightened F_0 might be a crucial factor for accent reduction in songs. F_0 is typically heightened in song, as well as extended over a broader range, in order to match the intended pitches within the composition's melody (or vocal harmony). Higher values of F_0 necessarily spread the harmonics of the resulting sounds further apart since all of the harmonics are integer multiples of the fundamental. Consequently, it is less likely that there will be harmonic energy near the intended formant center frequencies in order to adequately convey each formant, and thus, each phoneme. This under-representation of (natural resonances across) the spectrum should certainly make it more difficult for listeners to assess whether vocal tract resonances align with those typically generated by native speakers/singers. A similar explanation has been offered for observed reductions in identification performance for

spoken vowels with increases in F_0 (Ryalls & Lieberman, 1982). This possible explanation also aligns with previous findings that non-native speech with more F_0 variation is perceived as more native-like (e.g., see Polyanskaya et al., 2017; also see van Maastricht et al., 2021).

It can be effectively argued that this correlation was largely due to a categorical difference between speech and song, as a corresponding decrease in accentedness ratings with increasing median F_0 was not consistently observed when analyses were restricted to either singing or reading alone within a given gender. However, a corresponding correlation was obtained for female readers, that is, the only condition where there was a sufficient range of median F_0 values to statistically permit observation of any potential correlation (which did not occur for male readers; see Figure 5). Likewise, only the female readers produced F_0 values that were high enough to spread adjacent harmonics sufficiently far apart to potentially miss the intended center frequencies of one or more formants.

Thus, while the primary differences that were observed, and that provided the focus of the current investigation, were restricted to overall differences between speech and song, it is clear that, at least under some conditions, more systematic relationships between accentedness and F_0 should be expected.

In conclusion, it is clear that vowel lengthening does not consistently reduce the perception of foreign accentedness in either speech or song, and furthermore, that heightened F_0 values predicted the observed accent reduction for the original song stimuli at least as well, if not better than, mean vowel durations. Future evaluations for song would benefit from systematic and orthogonal (bi-directional) manipulations of both rate and F_0 parameters across a broad range of naturally acceptable values in order to determine their relative contributions to, as well as their possible perceptual interaction in, foreign accent reduction. Such evaluations will likely

either require that the singers initially produce corresponding samples separately across several pitch registers (e.g., by transposing melodies or utilizing different octaves) or, alternatively, a re-synthesis method that generally preserves the amplitudes of particular frequencies as F_0 is altered. Then a more complete understanding of the elements responsible for foreign accent reduction, at least for singing in English, should be attained.

Declaration of Interest Statement

The authors report there are no competing interests to declare.

References

Adobe Audition (Version 3.0). (2007). San Jose, California: Adobe Systems Incorporated.

Austin, S. F. (2007). Jaw Opening in Novice and Experienced Classically Trained Singers. *Journal of Voice, 21*(1), 72-79. doi: https://doi.org/10.1016/j.jvoice.2005.08.013

Beal, J. (2009). "You're Not from New York City, You're from Rotherham". *Journal of English Linguistics - J ENGL LINGUIST, 37*, 223-240. doi: 10.1177/0075424209340014

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer (Version 6.1.51). Retrieved from http://www.praat.org/

Chan, K. Y., Hall, M. D., & Assgari, A. A. (2017). The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology, 29*(1), 23-34. doi: 10.1080/20445911.2016.1170746

Christiner, M., & Reiterer, S. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Front Psychol, 4*. doi: 10.3389/fpsyg.2013.00874

Coumel, M., Christiner, M., & Reiterer, S. M. (2019). Second Language Accent Faking Ability Depends on Musical Abilities, Not on Working Memory. *Front Psychol, 10*. doi: 10.3389/fpsyg.2019.00257

Duan, Z., Fang, H., Li, B., Sim, K., & Wang, Y. (2013). *The NUS sung and spoken lyrics corpus: A quantitative comparison of singing and speech*.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America, 93*(3), 1589-1608. doi: 10.1121/1.406818

Gibson, A. (2010). *Production and perception of vowels in New Zealand popular music.* Auckland University of Technology.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21*(2), 205-228. doi: 10.1017/S0142716400002034

Hagen, M., Kerkhoff, J., & Gussenhoven, C. (2011). *Singing your accent away, and why it works*. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics, 16*(2), 153-169. doi: https://doi.org/10.1016/S0095-4470(19)30480-2

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417.

Mageau, M., Mekik, C., Sokalski, A., & Toivonen, I. (2019). Detecting Foreign Accents in Song. *Phonetica, 76*(6), 429-447. doi: 10.1159/000500187

Morrissey, F. (2008). Liverpool to Louisiana in one lyrical line: Style choice in British rock, pop and and folk singing. In L. M & S. a. J (Eds.), *Standards and Norms in the English Language: Contributions to the sociology of language*. De Gruyter, Berlin.

Munro, M. J. (1993). Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings. *Language and Speech, 36*(1), 39-66.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48*(2), 159-182.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition, 23*(4), 451-468. doi: 10.1017/S0272263101004016

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693-703.

Polyanskaya, L., & Ordin, M. (2019). The effect of speech rhythm and speaking rate on assessment of pronunciation in a second language. *Applied Psycholinguistics, 40*(03), 795-819. doi: 10.1017/s0142716419000067

Polyanskaya, L., Ordin, M., & Busa, M. G. (2017). Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. *Language and Speech, 60*(3), 333-355. doi: 10.1177/0023830916648720

QuestionPro. (2021). QuestionPro Survey Software. from https://www.questionpro.com/

Reiterer, S., Hu, X., Sumathi, T., & Singh, N. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Front Psychol, 4*. doi: 10.3389/fpsyg.2013.00782

Riazantseva, A. (2001). Second Language Proficiency and Pausing: A Study of Russian Speakers of English. *Studies in Second Language Acquisition, 23*(4), 497-526.

Ryalls, J. H. & Lieberman, P., (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America 72*(5), 1631-1634. https://doi.org/10.1121/1.388499

Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics, 37*(2), 303-322. doi: 10.1017/S0142716414000575

Simpson, P. (1999). Language, culture and identity: With (another) look at accents in pop and rock singing. *18*(4), 343-368. doi: doi:10.1515/mult.1999.18.4.343

Sundberg, J., & Skoog, J. (1997). Dependence of jaw opening on pitch and vowel in singers. *Journal of Voice, 11*(3), 301-306. doi: https://doi.org/10.1016/S0892-1997(97)80008-2

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25*(1), 1-24.

Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (2015). *Vowel height and duration.* Paper presented at the Proceedings of the 32nd west coast conference on formal linguistics.

Trudgill, P. (1983). *On dialect: Social and geographical perspectives*: Wiley-Blackwell.

van Maastricht, L., Zee, T., Krahmer, E., & Swerts, M. (2021). The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Communication, 133*, 81-90. doi: 10.1016/j.specom.2020.04.003

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of Acoustic Variability in the Perceptual Learning of Non-Native-Accented Speech Sounds. *Phonetica, 64*, 2-3.

Appendix A

The Participant Language Background Questionnaire

We are asking the following questions to determine if there are any circumstances that may influence how you perceive or produce certain sounds.

Age (in years)

Biological sex
1. Male
2. Female
3. Other (please specify) _____

Education Level (current)
1. Elementary School
2. High School
3. Some College or Technical Training
4. Bachelors
5. Masters
6. PHD

City & State of Birth (list also the country if it is not USA)

Is the City and State you lived in for the longest amount of time the same as your City and State of Birth?
1. Yes
2. No

Is English the first language you learned to speak?
1. Yes
2. No

Do you have either normal vision or corrected-to-normal vision (i.e., through corrective eyewear like glasses or contact lenses)?
1. Yes
2. No

Have you ever had a hearing or speech disorder?
1. Yes
2. No

Do you speak another language?
1. Yes
2. No

Do you speak that language fluently?

Do you have family members or close friends with a foreign accent?
1. Yes
2. No

Have you visited or lived in foreign counties that speak language other than English for more than 3 months?

1. Yes
2. No

Do you have regular contact with non-native speakers of English?

1. Yes
2. No

If you have regular contact with non-native speakers of English, please describe how often:

1. every day
2. every week
3. every two weeks
4. every month
5. every three months

Table 1

Means and Standard Deviations for the Durational Measures for Read and Sung Stimuli for each Self-identified Gender, as well as the Corresponding F-Statistics for Significant Differences Between Read and Sung Stimuli

	Read			Sung			F statistics for Significant Difference between Sung and Read Stimuli
Duration Measures	Female	Male	Total	Female	Male	Total	
Stimulus Duration	3.04 (.901)	3.32 (1.17)	3.18 (1.03)	6.29 (1.93)	6.24 (1.95)	6.26 (1.90)	$F(1, 44) = 47$, $p < .001$
Total Vowel Duration per Production (s)	1.30 (.467)	1.30 (.338)	1.30 (.399)	3.85 (1.39)	3.73 (1.27)	3.79 (1.30)	$F(1, 44) = 77$, $p < .001$
Mean Vowel Duration per Production (ms)	113 (26.4)	123 (45.8)	118 (36.9)	369 (194)	361 (153)	365 (171)	$F(1, 44) = 46$, $p < .001$
Non-vowel Duration	1.75 (.450)	2.03 (.944)	1.89 (.751)	2.44 (.959)	2.51 (.951)	2.47 (.935)	$F(1, 44) = 5.57$, $p = .023$
Proportion of Vowel in the Original Stimulus	.424 (.071)	.408 (.105)	.416 (.088)	.607 (.092)	.596 (.0852)	.602 (.0871)	$F(1, 44) = 51.7$, $p < .001$
Durational Ratio of Vowels to Consonants	.761 (.230)	.769 (.491)	.765 (.375)	1.75 (.992)	1.58 (.515)	1.66 (.778)	$F(1, 44) = 25$, $p < .001$

Note: All other main effects and interactions were not significant.

Figure 1

Mean Ratings of Foreign Accentedness (and Corresponding Standard Errors) as a Function of Production Rate Condition for Read and Sung Excerpts

Figure 2

Scatterplot of Mean Accentedness Rating and Mean Duration of Vowels (ms) within Excerpts Produced at the Original Rate by Female Speakers (filled circles), Female Singers (unfilled circles), Male Speakers (filled squares), and Male Singers (unfilled squares)

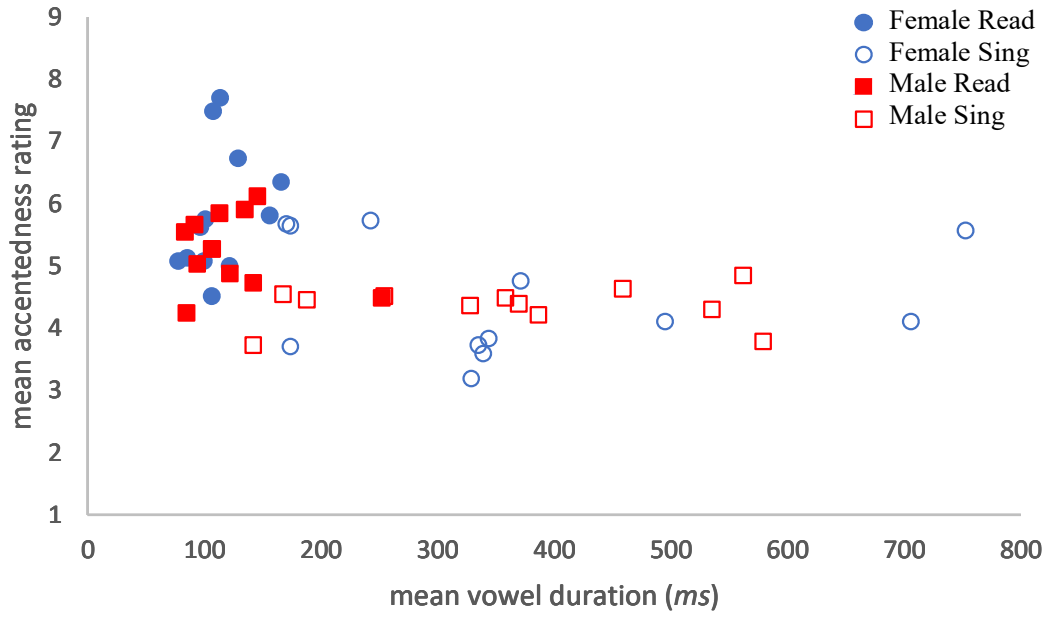

Figure 3

Scatterplot of Mean Accentedness Rating and Durational Ratio of Vowels to Consonants within Excerpts read and sung at the Original Rate by Female producers (unfilled circles) and Male producers (unfilled triangles)

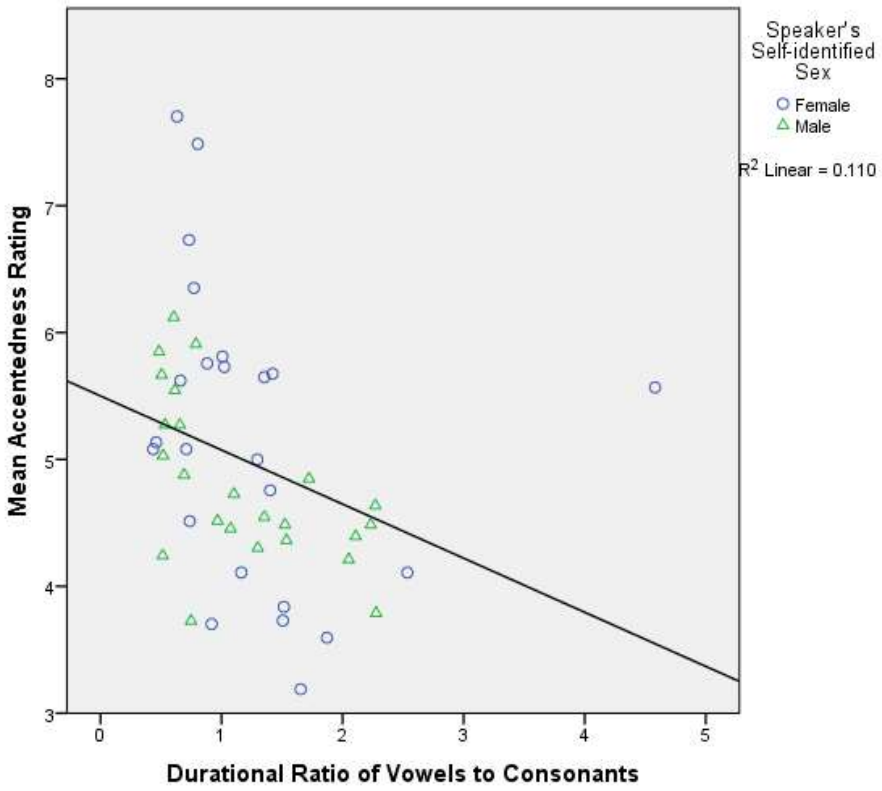

Figure 4

Scatterplot of Median F_0 Values (Hz) for Speech and Songs Produced by Male and Female Singers/Speakers at the Original Rate with the Speech-Song Equivalence Reference Line

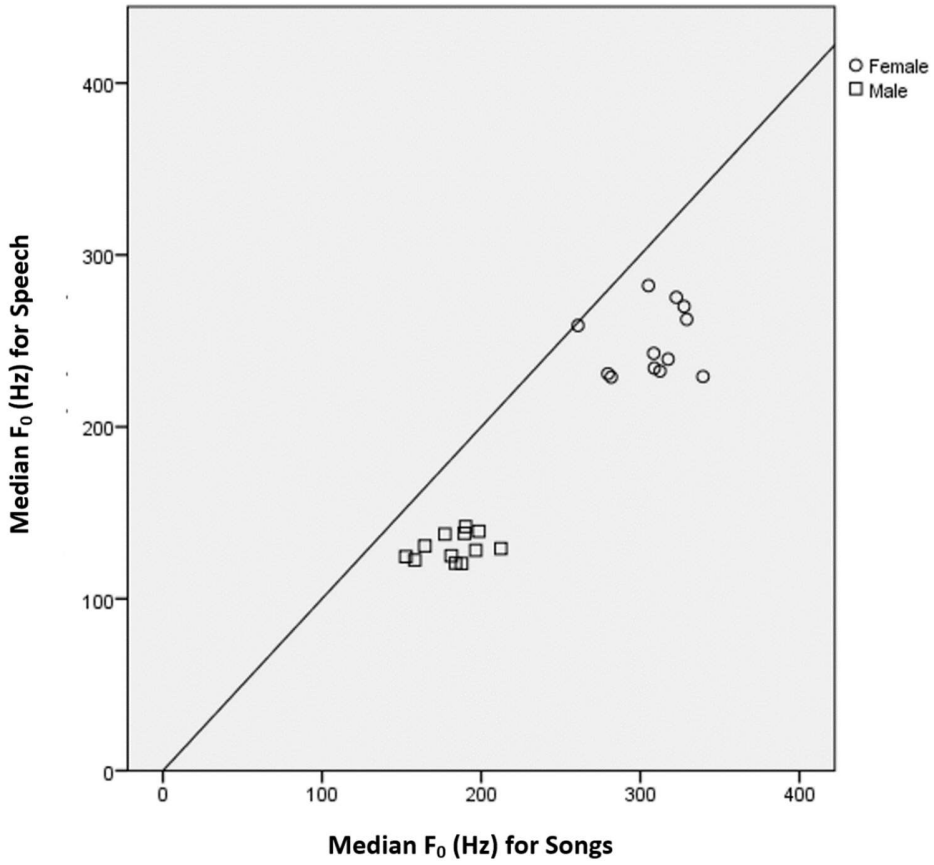

Figure 5

Scatterplot of Mean Accentedness Rating and Median F_0 Values (Hz) for Excerpts Produced at the Original Rate by Female Speakers (filled circles), Female Singers (unfilled circles), Male Speakers (filled squares), and Male Singers (unfilled squares)