Why Can You Sing Away Your Accent?

By: Ashley Sokalski

A thesis submitted in fulfillment of the requirement for CGSC 4908 as credit towards the degree

Bachelors of Cognitive Science with Honours

Specialization in: Biological Foundations of Behaviour

Supervised by: Dr. Ida Toivonen

Institute of Cognitive Science
Carleton University
Ottawa, Canada

April (2017)

Abstract

It is harder to perceive accents in song than in speech. This study proposes a possible mechanism through which accents are masked: minimization of the voiced-voiceless distinction that native English speakers maintain in speech. This was tested in a group of untrained singers with seven native and seven non-native English speakers. In the speech portion of the study, speakers were asked to read the words *bus, buzz, bet* and *bed*. In the singing version of the study, speakers were asked to sing each word to the tune of *Twinkle Twinkle Little Star*. Both native and non-native speakers maintained the voiced-voiceless distinction in song. However, the difference in vowel duration before voiced-voiceless consonants was greater in speech in native speakers than in non-native speakers, but in song that difference was eliminated. This difference between speech and song in native speakers is one potential reason why it is more difficult to detect accents in song than in speech.

Table of Contents

Introduction

It is significantly harder to perceive accents in song rather than in speech (Gibson, 2010; Hagen, Kerkhoff, & Gussenhoven, 2011; Mageau, 2016; Simpson, 1999; Trudgill, 1983). Most of the previous work on this topic has come from analyzing professional singers. There are currently two different hypotheses as to why accents are less noticeable in song. The first says that singers shift their accents according to the genre with which they hope to identify. The second hypothesis is that in song, syllable durations are imposed by rhythm and melody of the musical score so foreign accents tend to disappear or be minimally apparent. These two hypotheses are not incompatible with each other and it may be the case that both are true; however, this study will focus on the second hypothesis.

This study takes a closer look at how the rhythm of music influences the durational patterns of vowels. I specifically focus on the durational patterns of the voiced-voiceless distinction of English. In English, vowels that precede voiced consonants are longer than vowels that precede voiceless consonants. For example, the /ʌ/ in *buzz* is longer than the /ʌ/ in *bus*, because /z/ is a voiced consonant and /s/ is not. It may be that in order to follow the melody of the song, singers' use less of said distinction, making non-native and native speakers' vowel length very similar in song despite clear differences in speech. This proposed mechanism has never been studied in song in the context of accent detection, and the idea that accents are harder to perceive in song has rarely been tested in non-trained singers overall. The purpose of this thesis is to study whether native speakers maintain their voiced-voiceless distinction in song and if non-native speakers behave similarly. This will be tested in a group of untrained singers.

First, however, both hypotheses will be discussed along with the previous work done with trained and untrained singers in regards to music and accents. Finally, I will defend why I chose

vowel length as a phonological marker of an accent for my study and I will briefly explain the motivation behind my study.

1.1 Culture and Genre-Based Accent Shifting

In song, singers commonly use accents which differ from the ones they use in their ordinary everyday speech (Coupland, 2011; Gibson, 2010; Mageau, 2016; Simpson, 1999). These changes in accents may be because certain genres are connected to certain accents. In southern English for example, many singers do not shift their accents because they sing country, and country is commonly associated with a southern accent (Kelly, 2012). By that same token, rock or pop music is commonly associated with an American accent. However, this idea does not account for everyone and is again a tendency. The most commonly discussed example of cultural patterns affecting accent shift concerns British singers attempting unconsciously or consciously to sound more American by modelling American linguistic behaviour: where they impose stereotypically American features into their song lyrics (Gibson, 2011; Simpson, 1999). The frequent tendency has been described with reference to five phonological principles by Trudgill (1983). Trudgill's characterization has come to be called The USA-5 model and it includes the tendency to pronounce the vowels in words like *dance* in a more unrounded fashion.

Trudgill (1983) proposes that because America has a dominant influence on modern popular music, foreigners' admiration for American linguistic patterns naturally leads to imitation. Overall, accent shifting is done in the hopes that music will then gain popularity by imitating what singers know to be already positively perceived by audiences. Not all singers shift their accents, however. As mentioned above, speakers of southern English (i.e. English spoken in the American south) have been found not to shift their accents when singing, even though the West Coast American accent is much more accepted in the music business (Kelly, 2012).

While this hypothesis that accent shifting is connected to culture and genre is important, and probably correct, it will not be perused further here. The present study specifically asks whether accents are more difficult to detect in song than in speech even when the singer is not attempting to imitate an accent associated with a particular genre.

1.2 The Influence of Rhythm and Melody in Accent Detection

Non-native speakers are heard as having an accent because they deviate from native norms on the phonetic dimensions of pitch, loudness, duration, and quality. Hagen et al., 2011 and Mageau, 2016 propose that the rhythm and melody of music impacts the intonational factors of speech (as realized by pitch, loudness and duration), and that is why foreign accents tend to disappear or be minimally apparent in song.

It has been shown that trained singers do have less noticeable accents in song than in speech. Hagen et al.'s (2011) participants, who had been in the school choir and were either native or non-native-Dutch speakers, spoke and sang English song lyrics. Unbiased listeners were asked to rate how native each recording sounded on a scale of 1 (very strong foreign accent) – 7 (native speaker of Dutch). The scores for all the sung stimuli were significantly higher than the scores for the spoken stimuli, meaning listeners considered the sung stimuli to sound more authentic, regardless of if it was sung by a native or non-native Dutch speaker (Hagen et al., 2011). These results support the hypothesis that intonational cues are overridden by the rhythm and melody of the musical score, explaining why foreign accents tend to disappear, or be minimally apparent in song. However, all these trends were found among professional singers. This leaves a gap in the research; there is little research as to whether untrained singers also have less noticeable accents in song. This needs to be studied so that we can know if untrained singers behave like trained singers. This thesis will attempt to fill this gap by studying untrained singers.

There are, however, two main reasons why most research on how and why accents disappear in song is done with professional singers. The first reason being that singers often project a specific style that might be associated with a particular genre and accent (Trudgill 1983; Simpson 1999; Gibson 2011). Think here of an opera singer, who can sing in Italian perfectly, yet has no ability to freely engage in conversation. The second is that musical expertise tends to heighten a person's ability to reproduce an accent orally. The reason a speaker sounds foreign has to do with the fact that they have trouble with consonant, vowels, and diphthong pronunciation, as well as tonal replication. A singer, however, is taught special skills such as control of the articulatory apparatuses such as the tongue, which help them overcome these shortcomings. To illustrate, Christiner & Reiterer (2015) had German singers and instrumentalists listen to Hindi sentences. Participants were then asked to try and imitate the Hindi speakers they had listened to, Hindi being chosen because German speakers have little contact with the Hindi language. They showed that vocalists were significantly better imitators than instrumentalists (Christiner & Reiterer, 2015). It was posited that singing training heightens perceptual abilities, so that singers are able to discriminate rhythmical and tonal changes in voices more accurately than the other groups. As such, because professional singers are better able to mimic and manipulate accents, which is why most previous accent research has used professional singers as their samples.

According to the above studies, it may be that the reason why accents disappear with trained singers is because they specifically aim for genres associated with particular accents, and singers are trained in pronunciation and accent imitation. This thesis specifically focuses on what it is about singing per se that masks accents. In other words, what is it about singing that makes accents harder to detect when lyrics are not associated with a particular genre or culture. I suggest that the melody masks important intonational features marked by vowel length. I will test this hypothesis using untrained singers using a genre-neutral song. Since I hypothesize that accents

are perceived in large part due to variations in pitch, loudness and duration, I expect to find similarities in pitch, loudness and duration when native and non-native speakers sing a stimulus, and differences when native and non-native speakers speak the same stimulus, and I expect this to be true even with untrained singers. This study specifically focuses on vowel duration.

1.3 Previous Work With Untrained Singers

Mageau (2016) recorded six native speakers of English, and six second-language speakers. None of the participants were trained singers. They were asked to sing "Twinkle Twinkle Little Star" and read a passage from "Goldilocks". English listeners had more difficulty detecting a foreign accent in the singing conditions and rated speakers as having less of a foreign accent in song than in speech. These results are similar to those of Hagen et al. (2011) indicating that untrained and trained singers seem to behave similarly in regard to the fact that their foreign accents are masked. These results also suggest that it is more difficult to detect a foreign accent in song compared to speech, even when the singers are not trained, and are also not influenced by an accent associated with a particular genre (Mageau, 2016). Overall these findings give credibility to the hypothesis that in song, syllable durations are imposed by rhythm and melody of the musical score, so foreign accents tend to be minimized (Hagen et al., 2011; Mageau, 2016). This is also one of the only studies to use non-trained singers as participants, where vowel length calculations were specifically looked at. However, because the stimuli were not specifically chosen to be ideal for a vowel comparison, stimuli between the speech and song conditions could not be reliably compared. By controlling which vowels are used by adding voiced-voiceless minimal pairs, reliably measuring vowel length becomes possible between the speech and song conditions in the present study.

1.4 Vowel Length as a Marker of an Accent

In my study, the focus is on the fact that suprasegmental cues are partially masked by the melody when a person is singing, where my study specifically focus on vowel length as a marker of a foreign accent. But first of all, is vowel length a reliable marker of an accent? A neurological disorder known as Foreign Accent Syndrome (FAS) has provided much insight into the markers of an accent. The syndrome causes a native speaker to be perceived as foreign when speaking their mother tongue (Jose, Read, & Miller, 2016). This motor speech disorder can result in impairments of suprasegmental features, such as sentence stress, which can in turn influence syntax and affective output. When comparing FAS speakers to foreign accented speakers, as well as controls and speakers with a variety of speech language disorders, the FAS and foreign speakers groups were rated very similarly and higher in their degree of foreignness (Jose et al., 2016). Even though FAS is a neurological disorder, the overwhelming majority of listeners just hear an accent. When asked what markers made a listener label a speaker as foreign, the frequency of syntactic level errors and omission of morphemes were the most commonly reported answers (Jose et al., 2016). Interestingly, vowel length was among the markers of an accent as it approached significance. Jose et al. (2016) posit that vowel length can be perceived by listeners and helps listeners determine nativity of a speaker. However, the findings of the study are not conclusive, as it was not solely looking at vowel length as a marker of accents, and the data all came from self-report scales.

Wang and van Heuven (2003; 2005) have specifically studied vowel length by comparing English vowels of persons who have the tense-lax or voiced-voiceless distinctions in their language to those who do not natively have tense-lax and voiced-voiceless distinctions in their language.

English and Dutch vowels are either tense or lax, where lax vowels tend to be short, and tense vowels tend to be long. Mandarin/Cantonese does not use the length of a vowel as a phonological feature. Dutch, Mandarin and English speakers were thus asked to speak English diphthongs for vowel length analysis. As expected, the native English speakers clearly show that their lax vowels are significantly shorter (m= 169 ms) than their tense vowels (m=225ms) (Wang & van Heuven, 2006). The Mandarin speakers had both their long and short vowels within duration ranges considered normal for American English. Unexpectedly it was the Dutch speakers who made their long vowels [such as /ae/] too short by American standards, and their short vowels [such as /u/] too long (Wang & van Heuven, 2006). Even though Mandarin is a language without a tense-lax subdivision for vowels, Mandarin speakers were more successful in keeping the English vowel subdivisions distinct than were Dutch speakers. Moreover, Mandarin accented English vowels were more difficult for native English listeners to identify as being spoken by a foreigner than were Dutch accented vowels (Wang & van Heuven, 2003, 2005). Dutch speakers being unable to produce English vowels of the proper length, proved to be a significant phonological marker used by listeners in accent detection. This shows that suprasegmental characteristics, in this case vowel length, can differ in perceivable yet subtle ways, which are crucial to a listener attempting accent recognition.

In this study, vowel length will be investigated as an accent marker because the voiced-voiceless distinction can be empirically measured. If the vowel duration marking the voiced-voiceless distinction were neutralized in song, this would support the hypothesis that durational patterns are imposed by the melody and score of the song making native and non-native vowels sound more similar.

1.5 The Present Study

Hagen et al. (2011) and Mageau (2016) have proposed that accents are harder to perceive in song, in part because syllable durations are imposed by the rhythm and melody of the musical score, so foreign accents tend to disappear or be minimally apparent (Hagen et al. 2011; Mageau, 2016). However, with the exception of the Mageau (2016) study, no studies have looked at untrained singers. The present study will further investigate how music influences the natural durational patterns of speech in untrained singers, by specifically measuring vowel length. This study hypothesizes that one way in which accents are masked is through the elimination of the durational cues to the voiced-voiceless distinction that native English speakers maintain. In English, vowels are longer preceding voiced consonants compared to vowels that precede voiceless consonants. This can be seen by the fact that the /ʌ/ in *buzz* is always longer than the /ʌ/ in *bus*, because /z/ is a voiced consonant where /s/ is not.

More specifically, the study will focus on determining if the duration distinctive pattern of vowels before voiced and voiceless consonants are neutralized in song, and whether this occurs in both native and non-native speaker samples. Having native and non-native groups of speakers speak and sing minimal pairs that differ in voicing will test this hypothesis. In the singing condition participants will sing the same minimal pairs to the tune of Twinkle Twinkle Little Star, as it is a song that is not inherently associated without a particular accent. If the voiced-voiceless distinction is neutralized in song such that the /ʌ/ in *buzz* and *bus* become more similar in length, then this supports that syllable and vowel duration are imposed by melody and musical properties. In this case native and non-native speech patterns will become more similar, as the native English speakers would lose their voiced-voiceless distinction, and the non-native speakers never possessed the distinction. If the voiced-voiceless distinction remains, such that the /ʌ/ in *buzz* remains significantly longer than that of *bus,* then my proposed mechanism will not be supported.

Methods

2.1 General Methods

This study set out to determine if native and non-native speakers differed in song and speech specifically in vowel duration, which was our specific marker of an accent. More specifically, the goals of the study were to determine if the voiced- voiceless consonant distinction made by native speakers was also seen in non-native speakers and whether said distinction was neutralized in the song condition or not. The present study's methods are largely based on the methods of Mageau (2016).

2.2 Participants

Fourteen speakers, all females between the ages of 18-24 years old, were recruited from Carleton University to participate in the study. Recruiting took place via the Institute of Cognitive Science SONA system.

Seven of the participants were native speakers of English and seven were non-native speakers of English. The native speakers had been exposed to English since birth and had little exposure to a second language. The non-native speakers had not been immersed in an English-environment until adulthood. However, all speakers had some English exposure and/or training while growing up. The second language speakers' first languages were Arabic (four speakers) and Mandarin or Cantonese (3 speakers).

Participation was voluntary, and participants received compensation in the form of a 1% grade increase in a first or second-year Cognitive Science course, which they were enrolled in at the time of their participation in the study.

2.3 Stimuli

The minimal pairs *bed/bet*, *buzz/bus* were used as stimuli for both the spoken and the sung portion of the study. Each pair contains a word ending in a voiced consonant (buzz, bed) and a word ending in a voiceless consonant (bus, bet).

In the singing condition, the words *bed, bet, buzz* and *bus* were sung to the melody of *Twinkle, Twinkle Little Star*. This song was chosen by Mageau (2016) because the song does not belong to a genre associated with a particular accent that could potentially bias or influence the way a participant sang the song. In Mageau's study, participants sang the regular lyrics of *Twinkle, Twinkle.* In this study, I chose specific words suitable for comparing vowel duration before voiced and voiceless consonants.

A brief language questionnaire (*Appendix 3)* was administered to all speakers before the beginning of the study in order to gain insight into the participants' linguistic background. It asked if speakers natively spoke English. Speakers were also asked to include all other languages they spoke and rate from 1 (barely) – 5 (native) how well they felt they spoke each subsequent language.

2.4 Equipment

A *Tascam HD-P2* portable solid-state recorder paired with both head and table mounted *Audio Technica* microphones (one microphone for each recording channel), were used to record all participant sessions.

The acoustic analysis software PRAAT (Boersma, 2001) allowed for all segmentation and vowel length analysis.

A metronome set to 120 beats per minute was used in order to ensure that the participants maintained an even pace when they read and sang the target words.

2.5 Procedure

Speakers filled out the informed-consent form (*Appendix 2*) and language questionnaire *(Appendix 3)* while the audio recording device and metronome were being set up.

Before the singing portion of the study, the participants were asked if they wanted to hear an audio recording of *Twinkle Twinkle Little Star,* without lyrics to refresh their memories on the tune/ melody of the song. This recording was also played whenever any speaker requested it during the recording session.

In the speech portion of the study, speakers were asked to read the words *bus, buzz, bet* and *bed* 42 times each with the help of a metronome set to 120 beats per minute for 30 seconds to keep their pace steady. Each word was said 42 times because that number corresponds to the number of beats in one verse of *Twinkle Twinkle Little Star,* I wanted to ensure that the two conditions were as similar to each other as possible.

In the singing portion of the study, speakers were asked to sing each of the words *bus, buzz, bet* and *bed to* the tune of the first verse of *Twinkle Twinkle Little Star* (four words x 1 verse = four verses total). A metronome was again set to 120 beats per minute so that speakers would maintain a steady pace while singing.

Condition order was alternated so that half of the speakers began their recording session with the speech portion of the study and half with the singing portion of the study.

Following completion of both portions of the study, or whenever a participant asked to have their data withdrawn from the study, which happened on two occasions, they received their academic compensation. In total, the study took around 25 minutes for each speaker to complete from the start of the language questionnaire to the end of recording.

The data was then imported into *PRAAT* so that vowel length could be extracted for both the speech and song conditions. Importantly because the last syllable of each line of Twinkle Twinkle Little Star is a half note instead of the usual quarter note, the last word for each line of the song condition was removed, as it would have an artificially long vowel (Boersma, 2001). The sound files were also checked for error, any word that differed from the word the participants were supposed to be repeating was removed from the analysis.

Results

Vowels are longer before voiced than voiceless consonants. This is true cross-linguistically but the difference is especially large in English. The aim here is to determine if the same tendency occurs in song or if aspects such as rhythm neutralize it, and if non-native speakers abide by the same tendency.

Tables 1 and 2 lists each speaker's mean vowel duration for each of the vowels used in the study across both the singing and speech conditions. From the descriptive data, it appears that all speakers, regardless of if English is their first language, have longer vowels preceding voiced consonants than voiceless consonants across both the speech and song condition. On the whole, however, vowels in the singing condition are all longer than those of the speech conditions, even though the voiced-voiceless distinction remains. Table 3, which averages vowel length for each word for native and non-native speakers across speech and song, confirms these two trends and also shows that native speakers vowels on the whole tend to be longer than non-native speakers' vowels.

	Speech				**Song**			
Participant	**Buzz**	**Bus**	**Bed**	**Bet**	**Buzz**	**Bus**	**Bed**	**Bet**
4	201	145	168	140	289	191	229	193
7	213	144	177	109	220	200	217	176
12	207	149	213	126	279	196	238	142
13	278	146	177	144	284	227	216	176
15	139	125	126	113	199	174	171	155
16	200	162	182	121	174	172	164	144
17	209	139	196	155	224	180	203	169

Table 1. Native speakers mean duration of each vowel type for all participants across both the speech and singing condition.

	Speech				Song			
Participant	**Buzz**	**Bus**	**Bed**	**Bet**	**Buzz**	**Bus**	**Bed**	**Bet**
00	155	147	169	154	213	222	235	211
5	102	82	133	119	122	133	148	116
6	144	123	139	132	206	189	186	165
8	178	161	154	145	219	212	186	199
11	140	133	200	116	208	187	260	174
14	178	137	187	125	300	247	336	294
18	257	173	172	139	222	159	138	134

Table 2. Non-native speakers mean duration of each vowel type for all participants across both the speech and singing condition.

	Speech				Song			
Speaker	**Buzz**	**Bus**	**Bed**	**Bet**	**Buzz**	**Bus**	**Bed**	**Bet**
Native	207	144	176	130	238	191	205	165
Non-Native	165	137	165	133	213	193	213	185

Table 3. Average vowel length for native and non-native speakers for each word across both speech and song.

3.1 Vowel Duration in Song Versus in Speech

Native speakers spoken vowels averaged 164 ms and their sung vowels 200 ms. The non-native speakers show a similar lengthening trend in song with their spoken vowels averaging 150 ms and their sung vowels 201 ms. To see if vowel duration was significantly lengthened overall in the singing condition, one-way ANOVAs based on speaker types (native or non-native), using modality (speech or song condition) as the categorizing factor were used. They revealed that for both native and non-native speakers there was a significant effect of modality on vowel duration, where vowels were significantly lengthened in the song compared to speech condition ($F(1,53)=19.34$, $p < 0.01$) and ($F(1,53)=19.46$, $p < 0.01$) respectively. Both native and non-native speakers' vowels are longer in the singing condition.

As my experimental design used a 2 (native versus non-native English speaker) x2 (singing versus speaking) x2 (voiced versus voiceless vowel words) x1 (vowel length) mixed factor design, possible interactions were analyzed for significance. Native and non-native data were thus combined in a two-way ANOVA, where song still had a significant effect on vowel duration ($F(1,108)=29.52$, $p< 0.01$) as found before with the one-way ANOVAs. However, the interaction of song*speaker did not reach significance ($F(1, 108)=0.92$, $p=0.33$). This is crucial as it means that the finding that vowels are longer in song does not depend on which speaker was being looked at. All speakers behaved the same way. This confirms the trend seen in Tables 1 and 2.

3.2 Vowel Duration in Native Versus Non-Native Speakers

As seen in Table 3 native speakers appear to have longer vowels than non-native speakers in just over half of the words. In order to test if native speakers in general have longer vowels than do non-native speakers, an ANOVA was conducted. It did not reach significance ($F(1,109)= 0.72$, $p= 0.38$). This may be because the voiced-voiceless distinction is more pronounced in native speakers, so separating out the vowels by voicing, like in Table 3, makes them appear longer overall when statistically they are not. It may also be perhaps that native speakers vary their vowel length more or less than non-native speakers in certain conditions, and that is why accents are perceived. The significance of such effects will be considered now.

3.3 Effects of Consonants on Vowel Duration

From the descriptive data, it appears that both native and non-native speakers behave similarly: they have the voiced-voiceless distinction. Native spoken vowels preceding voiced consonants averaged 192 ms and vowels preceding voiceless consonants averaged 137 ms. Similarly, in song, vowels preceding voiced consonants averaged 222 ms and vowels preceding

voiceless consonants averaged 178 ms. Non-native speakers followed a similar trend, spoken vowels preceding voiced consonants averaged 165 ms and vowels preceding voiceless consonants averaged 135 ms. Similarly, in song, vowels preceding voiced consonants averaged 213 ms and vowels preceding voiceless consonants averaged 189 ms. Two-way ANOVAs separated by speaker type using voicing (consonant voicing) as the categorizing factor revealed that, for both native and non-native speakers, there was a significant main effect of voicing on vowel duration across both the speech and song condition ($F(1,54)= 27.6$, $p <.01$) and ($F(1,54)=4.103$, $p <.05$) respectively. Vowels are significantly longer before voiced consonants. This confirms that both sets of speakers maintained the voiced- voiceless distinction in both the speech and singing conditions. This shows that the aspects of rhythm and intonation, which are dictated by the song, do not fully neutralize this basic phonological tendency of English.

Based on the finding that both native and non-native speakers maintain their voiced-voiceless distinction, it was important to analyze possible speaker*voicing interactions. Both native and non-native voicing*speaker interactions did not reach significance ($F(1,52)=0.52$, $p=0.47$) and ($F(1,52)=1.33$, $p=0.25$) respectively. This is also an important finding, as it shows that maintaining the voiced-voiceless distinction in vowel length which is typical in English, did not depend on which speaker was being analyzed in both the native and non-native categories. All speakers behaved the same way. This also confirms the trend seen in Tables 1 and 2. When the interactions were combined into a voicing*song*speaker interaction, this as well was found to not be significant when native and non-native speakers were combined ($F(1,48)=0.037$, $p=0.85$). Simply put, vowel duration did not somehow depend on any specific combination of if the vowel was in the speech or song condition, and which speaker had uttered the vowel.

Of particular interest to my study as well is the fact that generally, the type of consonant that follows a vowel can also impact its duration. Generally vowels are longer preceding

fricatives, as in *buzz* versus stops, as in *bed* (Raphael, 1971). An ANOVA combining native and non-native speakers looking at the effect of consonant type and voicing on vowel length, found that voicing but not consonant type had a significant lengthening effect ($F(1, 109)=22.16$, $p <0.01$) and ($F(1,109=3.17$, $p=0.077$). It is important to note, however, that the effect of consonant type did approach significance, so this finding does not explicitly oppose the literature stating that vowels are longer following fricatives over stops. The interaction of voicing*fricative also did not reach significance ($F(1, 108)=0.02$, $p=0.87$). As both speaker types were combined in this ANOVA, the effect of speaker type and consonant type on vowel duration was analyzed, these also did not reach significance ($F(1,108)=3.16$,$p=0.08$) and ($F(1,108)=0.74$, $p=0.39$). This confirms that in both samples of speakers, using fricatives or stops does not affect vowel length significantly.

3.4 A Closer Look at Vowel Duration before Voiced & Voiceless Consonants

Descriptively it looks like native speakers make a larger voiced-voiceless distinction than non-native speakers in speech. A one-way ANOVA based on the duration, separated by speaker type (native or non-native), did not reach significance but approached it ($F(1, 108)=1.58$, $p=0.21$). However, vowels were overall longer in song, so it is possible that this distorts the perceived larger voiced-voiceless distinction in the speech of native speakers. Proportions of vowel length in both the speech and song condition were thus calculated to tease apart the length differences and test their significance. The voiced and voiceless word from each minimal pair was summed, and the voiceless word was subtracted out to obtain the proportion of the voiced minimal pair's vowel length in each minimal pair, in each speaker set across both study conditions.

Based on summary Table 4, it seems native speakers have a larger difference in vowel duration before voiced and voiceless consonants than non-native speakers in speech across both sets of minimal pairs. It also appears that less variation is seen in the song conditions for both minimal pairs across both speaker types. Vowels are more consistent or similar in song in both speaker samples.

Minimal Pair	Speaker Type	Voicing	Speech	Song	Proportion
	Native	0	144	191	-
		1	207	238	-
	Native Total	-	**351**	**429**	**59% vs 56%**
Buzz/Bus	Non-Native	0	137	192	-
		1	165	213	-
	Non-Native Total	-	**302**	**405**	**55% vs 53%**
	Native	0	130	165	-
		1	177	205	-
	Native Total	-	**307**	**370**	**58% vs 55%**
Bed/Bet	Non-Native	0	133	185	-
		1	165	213	-
	Non-Native Total	-	**298**	**398**	**55% vs 54%**

Table 4. Mean vowel length durations of each minimal pair (buzz/ bus & bed/bet) for native and non-native speakers across speech and song conditions. The total row sums the voiced and voiceless durations for each of the modality conditions. Proportions of vowel length for speech and song conditions, respectively, are in the last column.

To analyze the significance of the overall proportions, each individual speaker's proportions were calculated for each minimal pair for both native and non-native speakers. A Welch two sample t-test comparing native and non-native vowel length proportions in song, found no significant difference at the $p<.05$ level. When comparing native and non-native vowel

length proportions in speech, however, there was a significant difference ($t(24)=2.17$, $p=0.04$, 2-tailed). Vowels preceding voiced consonants have vowel durations, which vary most in speech for native speakers. Native speakers make a greater vowel distinction in speech than in song.

A Welch two sample t-test comparing non-native song and non-native speech found no significant difference in the proportion of vowel duration at the $p<.05$ level. When comparing native speech and song, however, there was a significant difference ($t(24) =-2.0874$, p-value = 0.048, 2-tailed). This confirms that, as seen in the summary proportion tables above (Table 4) the proportional length of vowels preceding voiced consonants are greater in speech than in song for native, but not for non-native speakers. This is a possible explanation for why it is harder to hear accents in song: native speakers make a greater vowel distinction in speech than in song. In other words, native speakers voiced-voiceless distinction is minimized in song making their vowel length more similar to non-native speakers vowel length, thus making it harder to perceive accents in song.

3.5 Results Summary

Overall, looking back at the goals of the study, we now have some answers. Confirming previous research, it was found that overall vowels are significantly longer in song. Related to the research hypothesis it was found that native speakers maintain the voiced-voiceless distinction even in song. Moreover, non-native speakers of English also maintain this characteristic distinction in song. The study did, however, add to our understanding of why it is more difficult to detect accents in song. Analyzing voiced vowel length proportions, it was shown that the vowel durations of native speakers differ more from the vowel durations of non-native speakers in speech than in song. And finally, the tendency that certain consonants, specifically fricatives, affect vowels that precede them is a trend in this study although it did not reach significance.

Discussion

Previous research has found that accents are harder to detect in song (Hagen et al., 2011; Mageau, 2016). This study built on the proposed view that the melody of the song partially masks suprasegmental cues. Specifically, I proposed that this is in part because of masking of the voiced-voiceless distinction, making native and non-native speakers vowels sound more similar. While most previous research has been done with professional singers, this study tested the hypothesis in untrained singers.

In line with previous research, vowels were lengthened in song for all speakers. In addition, the study evaluated vowel length preceding stops versus fricatives, as research has shown that there is a trend for vowels to be longer preceding fricatives. While the effect was not significant here, the results did trends in support of the literature. While the voiced-voiceless distinction is a very robust finding cross-linguistically, the idea of consonant types affecting vowel length is not as robust, and as such it is not surprising that this result did not reach significance.

Interestingly, and somewhat contrary to the hypothesis of the study, both native and non-native speakers maintained the voiced-voiceless distinction in song.

To investigate this finding further, proportions of vowel lengths preceding voiced and voiceless consonants were calculated for each speaker type for each set of minimal pairs (*bus/ buzz* and *bet/ bet*). Here it was found that proportions of vowel length vary most in native speakers speech and that in song this variation is smaller. In other words, native speakers make a greater voiced-voiceless distinction in speech than in song. This is a possible explanation for why it is harder to hear accents in song: even though both speaker sets maintain their voiced-voiceless distinction, some of the variation in vowel length characteristic of native speakers is erased

making them sound more similar to the non-native speakers. While the hypothesized mechanism of song masking the voiced-voiceless distinction was not completely supported, this study suggests a different mechanism through which song masks suprasegmental cues; by minimizing the vowel length variations characteristic of native speech, and slightly reducing their voiced-voiceless distinction. Overall, this study contributes to our understanding of how the rhythm and melody of song affects pronunciation, and adds to the research on accents in song more generally.

4.1 Limitations and Future Research

This was one the first studies using untrained singers, besides Mageau (2016), to potentially find a new way in which song masks intonation/suprasegmental cues. However as with any study, there are limitations to this finding. The sample size was small, with seven speakers in each condition, for a total of fourteen speakers. With a sample this small, statistical power is low when running statistical tests on the results. This means that the study may not have been able to detect significant differences in cases where there may have been one. However, this means that where significant differences were found, for example in the maintenance of the voiced-voiceless distinction in song and the loss of native vowel length variability in song, these results are even more powerful. With low power, these differences were still detectable, indicating the magnitude of their effects. Importantly though, because of the small sample size, it is important that this study be replicated in order to determine if the voiced-voiceless distinction is maintained in song for the general population, and if so, if native speaker vowel length variability is also erased.

On the topic of replication, it will also be important that future studies test this phonological mechanism of accent masking in both males and females, as this study focused solely on females. While this was a deliberate decision as males and females voices differ

occasionally especially in regard to pitch, as females have a higher pitch on average in everyday speech, it is important that the results be generalized across speaker populations.

While generalizability across the sexes is important, so is generalizability across second languages. In order to recruit enough participants for the study, I did not control for the second languages that the seven non-native English speakers spoke. Most of the non-native speakers spoke Arabic, which is a language that also has the voiced-voiceless distinction, which is also characteristic of English. This may have biased the results and made the native and non-native vowels quite similar from the beginning. For future research, the optimal study would have a group of native English speakers, a group of non-native speakers whose second language also has the voiced-voiceless distinction, such as French or Dutch and a group of speakers whose second language has no such distinction such as Mandarin or Cantonese. Then these three groups would be compared to see if one group's voiced-voiceless distinction in English was neutralized more in song than the other groups. At the same time, minimization of variation in their vowel length durations in song, a possible mechanism of accent masking as proposed by this study could also be analyzed.

And lastly, a methodological limitation was that the metronome used ended up coming through quite heavily on the recordings of the participants. While this did not directly affect vowel length calculations, it did mean that speakers F1 and F2 formants could not be reliably extracted and used in this study. As such, future research should look for a way around this sound issue, perhaps by giving participants headphones through which they hear the metronome so it does not interfere with recording. This is an important area of future research as it is very possible that native and non-native speakers have significantly different formants, whose variation may again be erased in song as was vowel length variation in this study, and this may

further support the hypothesis that melodies partially mask suprasegmental cues making accents harder to perceive in song. As such future research is needed.

Works Cited

Boersma, P. (2001). Praat, A System for Doing Phonetics by Computer. *Glot International*, *5*(10), 341–345.

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers Outperform Instrumentalists on Foreign Accent Imitation. *Frontiers in Human Neuroscience*, *9*(8), 482–490. http://doi.org/10.3389/fnhum.2015.00482

Coupland, N. (2011). Voice, Place and Genre in Popular Song Performance. *Journal of Sociolinguistics*, *15*(5), 573–602. http://doi.org/10.1111/j.1467-9841.2011.00514.x

Gibson, A. (2010). Production and Perception of Vowels in New Zealand Popular Music.

Gibson, A. (2011). Flight of the Conchords: Recontextualizing the Voices of Popular Culture. *Journal of Sociolinguistics*, *15*(5), 603–626. http://doi.org/10.1111/j.1467-9841.2011.00515.x

Hagen, M., Kerkhoff, J., & Gussenhoven, C. (2011). Singing Your Accent Away, and Why it Works. *17th International Congress of Phonetic Sciences (ICPhS XVII)*, (8), 799–802.

Jose, L., Read, J., & Miller, N. (2016). Is Language a Factor in the Perception of Foreign Accent Syndrome? *Language and Speech*, *59*(2), 219–235. http://doi.org/10.1177/0023830915586612

Kelly, N. (2012). *Production and Perception of Intonation in Southern American English and West Coast American English*. *Regional Variation in Intonation* (Doctoral Dissertation). University of Texas at Austin.

Mageau, M. (2016). *Foreign Accents in Song and Speech* (Master's Thesis). Carleton University.

Raphael, L. (1971). Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *The Journal of the*

Acoustical Society of America, *51*(4), 1296–1303.

Simpson, P. (1999). Language, Culture and Identity: With (another) Look at Accents in Pop and Rock Singing. *Multilingua*, *18*(4), 343–68. http://doi.org/10.1515/mult.1999.18.4.343

Trudgill, P. (1983). *On Dialect: Social and Geographical Perspecticves* (1st ed.). Basil Blackwell Publishing Ltd.

Wang, H., & van Heuven, V. J. (2003). *Mutual Intelligibility of Chinese, Dutch and American Speakers of English*. *Linguistics in the Netherlands*. Amsterdam & Philadelphia: John Benjamins.

Wang, H., & van Heuven, V. J. (2005). Mutual Intelligibility of American, Chinese and Dutch-accented Speakers of English. In *Proceedings of Interspeech* (pp. 1101–1104).

Wang, H., & van Heuven, V. J. (2006). Acoustical Analysis of English Vowels Produced by Chinese, Dutch and American Speakers. *Linguistics in the Netherlands*, *23*(1), 237–248. http://doi.org/10.1075/avt.23.23wan

Appendix

Appendix 1

Dear Sir or Madam,

We are writing to you today to invite you to participate in a study entitled "Music, accents and vowel length". We are Ashley Sokalski and Ida Toivonen at Carleton University. This study explores native and foreign accents in song and speech.

You will be recorded singing and reading, "Twinkle Twinkle Little Star". You will also be asked to sing and read the words bus, buzz, bet, bed 42 times each . The entire session will take between 15 and 30 minutes.

You will have the right to end your participation in the study at any time. Any incomplete recordings will not be analyzed and the data will be destroyed.

This project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. (Clearance expires on: August 31st 2017) Should you have questions or concerns related to your involvement in this research, please contact the REB Chair, Professor Andy Adler at ethics@carleton.ca or at 613-520-2517.

If you would like to participate in this research project please contact Ashley Sokalski at AshleySokalski@cmail.carleton.ca or Ida Toivonen at ida.toivonen@carleton.ca

Sincerely,

Ashley Sokalski and Ida Toivonen

Appendix 2

Carleton UNIVERSITY
Canada's Capital University

Consent of the participant

Title: **Music, accent and vowel length**

Date of ethics clearance: June 17th 2016

Ethics Clearance for the Collection of Data Expires: August 31st 2017

I, _____, choose to participate in a study on accents in song and speech. My participation will consist of reading the words *bus, buzz, bet* and *bed* 42 times each. As well as singing *bus, buzz, bet* and *bed* to the tune of *Twinkle Twinkle Little Star.*

We will record this and analyze the recordings phonetically. This project is carried out by Ashley Sokalski under the supervision of Dr. Ida Toivonen. Ashley Sokalski is an undergraduate student in Cognitive Science at Carleton University. Ida Toivonen is Associate Professor in Cognitive Science and Linguistics at Carleton University.

The participant has the right to end his or her participation in the study at any time, up to the completion of the recorded oral exercise. Participants who withdraw from the study will have their information immediately destroyed. Due to the promise of anonymity, it is not possible to withdraw after the recording is complete.

Participants will remain anonymous in this study. Names and other personal identifiers will not be attributed to the observations noted. The data will be stored on the researchers' password protected computers. The recordings will be stored securely upon completion of the research project. **The data might be saved and used in future studies.**

This project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. Should you have questions or concerns related to your involvement in this research, please contact:

Professor Andy Adler, Chair
Research Ethics Board
Carleton University
1301 Dunton Tower
1125 Colonel By Drive

Ottawa, ON K1S 5B6
Tel: 613-520-2600, ext. 3591w

Researcher contact information:
Ashley Sokalski
Institute of Cognitive Science
Carleton University

Email: ashleysokalski@cmail.carleton.ca

Supervisor contact information:
Ida Toivonen
Institute of Cognitive Science
Carleton University
Tel: 613-520-2600, 1202
Email: Ida_Toivonen@carleton.ca

By signing this, it will be understood that you have consented to participate in the project, and that you consent to the publication of the results with the understanding that anonymity will be preserved. If you agree to participate in this study and are over the age of 18, please sign below.

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

Appendix 3

List each language you know, and give an approximate rating of your knowledge of each language:

1) _____

native, near native, fluent, almost fluent, basic

2) _____

native, near native, fluent, almost fluent, basic

3) _____

native, near native, fluent, almost fluent, basic

4) _____

native, near native, fluent, almost fluent, basic

5) _____

native, near native, fluent, almost fluent, basic

6) _____

native, near native, fluent, almost fluent, basic