Speech, Language and Hearing

ISSN: 2050-571X (Print) 2050-5728 (Online) Journal homepage: http://www.tandfonline.com/loi/yslh20

The bilingual voice: Vocal characteristics when speaking two languages across speech tasks

Binna Lee & Diana Van Lancker Sidtis

To cite this article: Binna Lee & Diana Van Lancker Sidtis (2017): The bilingual voice: Vocal characteristics when speaking two languages across speech tasks, Speech, Language and Hearing, DOI: 10.1080/2050571X.2016.1273572

To link to this article: http://dx.doi.org/10.1080/2050571X.2016.1273572

The bilingual voice: Vocal characteristics when speaking two languages across speech tasks

Binna Lee [1,2], **Diana Van Lancker Sidtis**[1,2]

[1]Department of Communicative Sciences & Disorders, New York University, New York, NY, USA, [2]Brain and Behavior Laboratory, Geriatrics Division, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

The present study compares vocal parameters in bilinguals' use of two languages across different speech tasks, in order to examine whether language selection and task type will manifest different outcomes. Korean-English (KE) and Mandarin-English (ME) speakers performed three speech tasks – a reading passage, a monologue, and a picture description – in their two spoken languages. Fundamental frequency (F0), F0 variability, intensity, and speaking rate were measured. Results yielded an effect of language on average F0s, as both bilingual groups showed a significantly higher F0 in Mandarin or Korean compared to English. For the KE group, a higher F0 appeared in Korean compared to English across all tasks, while for the ME group, a difference in F0 between Mandarin and English was found for reading only. Both groups had the highest F0s in the reading task. There was a difference in F0 variability between two spoken languages only for the KE group monologue. Mean intensity of the ME group was overall higher than that of the KE group. Speaking rate was significantly slower in English than the native language for both groups, but the KE group demonstrated a greater difference in pace between their two spoken languages. Findings demonstrated that bilingual speakers produced notably different voice patterns contingent on language and speech task, indicating that inter- and intra-speaker variability in speakers' vocal features can be attributed in part to language effects.

Keywords: Voice, Task effect, Fundamental frequency, Bilingualism, Korean, Mandarin

Introduction

According to the US census Bureau and 2011 American Community Survey, more than 350 different languages are spoken in US homes. In great metropolitan areas such as New York or Los Angeles, approximately 38 –54% of residents are reported to speak a language other than English at home. These facts reveal increasing diversity of language groups in the US followed by enhanced awareness of the impact of these spoken languages on aspects of individuals' performances. The current study expands this line of interest on diversity of spoken languages through an investigation of voice profiles of bilingual speakers and examines the vocal influence of language selection and speech task on their speech performance.

The term 'voice' has many definitions, depending on context and perspective. Kreiman and Sidtis (2011) distinguish between voice narrowly considered, referencing laryngeal-supraglottal phonation, and voice broadly considered, which is synonymous with speech. The voice pattern broadly considered manifests auditory-acoustic parameters such as fundamental frequency mean and variation, intensity mean and variation, temporal cues including pausing and rate, voice quality, such as harsh or breathy, and articulatory features of pronunciation, all of which listeners use to make judgments about the speaker's personal characteristics and intentions. In this study, 'voice' refers the vocal pattern broadly considered, with a focus on fundamental frequency (F0), intensity, and rate. The other variable in this study, the bilingual, can also benefit from a definition. The term bilingualism is defined here as the ability to use two different languages with a range of linguistic abilities (speaking, reading, and writing) in different social contexts (Roberts and Shanker, 2007).

Over the past few decades, a number of voice studies have undertaken empirical investigations to explore the effect of language on the acoustic aspects of speech focusing on a variety of different language groups (Andrianopoulos *et al.*, 2001; Altenberg and Ferrand, 2006; Awan and Mueller, 1996; Hudson and Holbrook, 1982; Jarvinen *et al.*, 2013; Mayo

Correspondence to: Binna Lee, Steinhardt, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, 9th Floor, New York, NY 10012 USA.
Email: bnlee@nyu.edu

DOI 10.1080/2050571X.2016.1273572

and Manning, 1994; Mennen *et al*., 2012; Morris, 1997; Ng *et al*., 2012; Sapienza 1997; van Bezooijen, 1995; Xue *et al*., 2001; Wheat and Hudson, 1988). In a majority of these comparative studies, F0 differences have been the focus of interest. In one of the earlier studies, for example, Hanley *et al*. (1966) compared acoustic values of F0 in speakers of Tagalog, Spanish, Japanese and English. The results indicated that the English group had lower F0 measures compared with Japanese and Spanish groups. In another study, Andrianopoulos *et al*. (2001) compared F0s in isolated vowel productions in speakers of English, Mandarin, and Hindi, and found that Mandarin speakers had significantly higher F0 values compared to the other speakers. Japanese and Dutch women also exhibited different acoustic speech patterns, wherein the former speakers used a higher range of F0 (van Bezooijen, 1995). Similar findings of F0 differences among other ethnic groups were observed; elderly Euro-American females had a significantly a higher mean F0 than African-American females (Xue *et al*., 2001). More recently, Keating and Kuo (2012) compared the F0 of Mandarin and English monolinguals and found that Mandarin speakers had a higher F0 and larger F0 range than English speakers. As shown, different language groups exhibited difference in F0 parameters.

However, other studies failed to find difference in F0s across various language and ethnic groups. For example, Sapienza (1997) reported no significant differences in mean F0s in African-American and Caucasian speakers when producing English vowels. Awan and Mueller (1996) further compared mean F0 measures in connected speech among African-American, Caucasian -American, and Hispanic kindergartners. In their findings, except for African-American children who had a significantly lower F0 measures than Hispanic children, the remaining groups revealed comparable F0s. Not only average F0s, but F0 variability revealed no differences among different language groups (Baken and Orlikoff, 2000; Gelfer and Denor, 2014).

Although these findings are inconsistent, some studies do reveal a relationship between language and vocal attributes. There has been a heavy focus on F0, and not all contributory vocal characteristics have been examined. It is noteworthy that intensity, to our knowledge, has not been compared across languages, although loudness is a strong communicator of personal states and traits, such as anger and aggression (Kreiman and Sidtis, 2011). It is likely that linguistic groups may differ on this parameter. Another strongly signaling vocal cue, rate, has received only a little attention in this domain of study, even though syllable rate differences between language types are impressionistically quite obvious (Amino and Osanai, 2015).

The studies reviewed above compared voices of monolingual speakers, so that the observed differences in vocal measures might be attributable to anatomical factors rather than to cross-linguistic differences (Ng *et al*., 2012; Todaka, 1993). To exclude the possibility of anatomical or morphological differences from ethnicities as a confounding factor, recent studies examined vocal features of an ethnically-uniform selection of bilingual speakers (e.g. Altenberg and Ferrand, 2006; Jarvinen *et al*., 2013; Ng *et al*., 2010; Ng *et al*., 2012; Todaka, 1993; Yamazawa and Hollien, 1992). Parallel to the monolingual voice studies, however, these bilingual studies failed to reach a consensus in their findings of language difference. Whereas some studies show promising results of the effect of language on vocal parameters, especially F0 (Jarvinen *et al*., 2013; Keating and Kuo, 2012; Ng *et al*., 2010; Ng *et al*., 2012; Todaka, 1993; Yamazawa and Hollien, 1992), others failed to show a difference between spoken languages in bilinguals (Altenberg and Ferrand, 2006). Challenges included methodological limitations (e.g. restricted selection of vocal measures; order effects), limited speech sample selection, and lack of consideration for second language factors.

With respect to factors that relate to the language effect on bilingual voice, two major underlying aspects in the use of different languages should be considered: the intrinsic linguistic features of each language, and the contingencies of foreign language learning. Both elements can affect the way the voice is used in different languages, and thus influence the measured outcomes of various vocal characteristics in bilingual speakers.

Regarding the linguistic characteristics of language, vocal aspects of speech can be influenced by components such as tonal features (Altenberg and Ferrand, 2006; Keating and Kuo, 2012; Ng *et al*., 2010; Ng *et al*., 2012) or tense consonants (Jun, 1993). Tone languages such as Mandarin and Cantonese include lexical tones that represent different F0 features, which can relate to a larger range in F0 and, possibly, a higher average F0 during connected speech. For example, Adrianopoulos *et al*. (2001) reported in their study that Mandarin speakers showed higher F0 and F0 ranges compared with Hindi and American-English speakers when producing English vowels. Similar findings were found in another study comparing Mandarin monolinguals with English monolinguals (Keating and Kuo, 2012). Additionally, tense consonants in Korean are understood to function like lexical tones, with different F0 characteristics (Jun, 1993). Therefore, the influence of consonants on F0 in Korean can be considered to

be tonal, as certain types of consonants predict a tonal pattern. For instance, Korean introduces intrinsic pitch features of tense and tone with the production of certain consonants. The effect of tense consonants on F0 in Korean is shown to spread across multiple syllables, which can be represented phonologically with tonal characteristics (Jun, 1993). Specifically, voiceless consonants and tense consonants correlate with high tonal features, which can affect the overall F0 contours. Thus, different vocal patterns could possibly be the result of the linguistic attributes of a language itself.

As for the effect of second language use, degree of language proficiency may have an impact on the oral production of bilingual speakers (Jarvinen *et al.*, 2013; Ohala, 1984). A number of factors are claimed to impact one's second language proficiency such as age of arrival (AoA) in a non-native-speaking country, length of residence, age of second language learning, amount of native language use, motivation, and others (Asher and Garcia, 1969; Bialystok, 1997; Flege and Fletcher, 1992; Flege *et al.*, 1995; MacKay et al., 2000; Oyama, 1976; Thompson, 1991).

Many of these variables are highly correlated to degree of perceived foreign accent in the second language (Asher and Garcia, 1969; Flege and Fletcher, 1992), which may involve various aspects of prosody, fluency, and articulation. Accent, slower speech rate, increased F0, and low intensity in the foreign language (Munro and Derwing, 2001) may evoke different judgments of bilingual speakers. Slow rate may evoke impressions of uncertainty and while high f0 may arise from psychological anxiety, contributing to tension in the laryngeal structures (Jarvinen *et al.*, 2013; Ohala, 1984). Soft speech can signal passivity, insecurity, or compliance. For more competent individuals, however, dual language learners may modify their speech characteristics to accommodate the second language, as a result of adaptation and second language-related experience (Jarvinen *et al.*, 2013). These findings bring more attention to vocal attributes in the bilingual population, which may be altered according to language selection and degree of second language competency.

The present investigation further highlights a noticeable limitation in previous studies, variability in the type of speech tasks, and attempts to address the issue. The influence of task type is not a novel concept in the literature on voice studies (Reich *et al.*, 1989; Zraick *et al.*, 2005). Selection of different speech materials can exert noticeable effects on vocal features within individuals: specifically, many studies have illustrated that an individual's performance level can vary, depending on whether a task is structured or unstructured, with the former producing more advantageous results in performance (Baker *et al.*, 2008; Hunter, 2009). Significantly different outcomes in acoustic vocal measures were reported to be highly task-oriented (Baker *et al.*, 2008; Hunter, 2009; Keating and Kuo, 2012; Mang, 2001; Zraick *et al.*, 2005). Specifically, average F0 was significantly higher in reading tasks than in spontaneous speech samples across various populations of children, young adults, older men, and adult trained singers (Hudson and Holbrook, 1982; Mysak, 1959; Ramig and Ringel, 1983; Sorenson, 1989). Also, it has been observed that spontaneous speech constitutes a more natural form of production in habitual voice than other tasks, leading to difference in vocal characteristics (Kreiman and Sidtis, 2011). Thus the selection of task has been known to affect a speaker's speech performance in many aspects. However, a majority of the bilingual voice studies restricted the conditions of participants' performance to a single task, and in most cases, spontaneous speech was not included in the design. The failure to obtain naturalistic speech data when exploring aspects of vocal production should not be disregarded. In order to capture the voice profile of individuals in the most naturalistic way, performance measures should be considered in the broader context of connected speech.

This study aims to resolve the aforementioned issues by investigating whether selection of language will manifest different outcomes of vocal characteristics across speech tasks in bilingual speakers. We compare four vocal parameters in female Korean-English and Mandarin-English bilingual individuals, who are proficient in both languages, across different speech tasks. Second language proficiency was evaluated through personal interview by a certified ESL teacher (BL), a language questionnaire, and a foreign accent rating task. We selected two different groups of bilingual speakers in order to study the effect of using two distinct languages within an individual. The measures under investigation for bilingual participants were (1) mean F0, (2) F0 variability, (3) intensity (loudness), and (4) rate (syllables per second).

Methods

Participants

A total of 22 female bilingual speakers were recruited to participate in the present study, consisting of 11 Korean-English speakers and 11 Mandarin-English speakers. The Korean-English bilinguals (KE group) had a mean age of 25.55 years (range = 21–27 years), and the Mandarin-English bilinguals (ME group) had a mean age of 23.36 years (range = 22–25 years). Participants' education levels ranged from baccalaureate to post-graduate. All of the selected participants met the following selection criteria for bilingualism: (a) their native language was either Korean or Mandarin; (b) their parents were non-native

speakers of English; (c) they attended or were attending a school in the United States; (d) they were fluent or proficient in English determined by an initial screening interview by the first author (BL, a fluent Korean-English bilingual, has teacher certification in 'teaching English as a second language'). Second language proficiency is defined in many terms and used differently by various researchers (for a review, see McNamara, 1996). Among the different components of competency in second language, oral proficiency was considered the crucial determining factor that reflects different patterns of vocal features in speech production. Thus, participants' English oral competency was determined by two conditions: an oral screening interview, and foreign accent rating in English. For the oral screening interview, participants were responded to questions in English such as 'How are you feeling today?' or 'How did you find out about the study?' Potential participants with nonfluent, inaccurately expressed, or ungrammatical responses were excluded from the study.

The KE group reported speaking the Seoul dialect of Korean, and Modern Standard Chinese (Mandarin, Putonghua) was spoken by the ME group. All participants attested that their voice sounded normal and that their health status was good or excellent on the day of testing. Written consent was obtained from all participants using the Human Subjects Consent Forms, as mandated and approved by the Internal Review Board of the University.

Recording procedure and tasks

Participants' speech was recorded in a sound-attenuated booth in the Speech and Hearing Clinic at New York University. At the end of each speech recording session, a written questionnaire was administered to participants to obtain information on their language and educational background. A description of the demographic information of the participants is displayed in Table 1.

All participants performed three different speech tasks: a reading passage, a monologue, and a picture description, each in their two spoken languages. All instructions were provided in written form. The research design consisted of two experimental blocks, wherein the use of only one language was permitted per block. To control for order effects, the order of language was counterbalanced across participants and within each language, and the order of tasks was randomized. In all, there were six possible presentation orders, combining the order of the language and task.

Reading task

A reading passage was provided to the participants in English and in their native language. Each participant was asked to read 'The Rainbow Passage' (Fairbanks, 1960) at a normal speech rate, using a comfortable speaking voice. For Mandarin-English speakers, a translated version of 'The Rainbow Passage' was provided in Mandarin. For Korean-English speakers, the same reading passage was translated in Korean. All participants were allowed to practice reading the passage a number of times to familiarize themselves with the task, before the speech samples were recorded.

Picture description

The 'Cookie Theft' picture of the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1972) was used to elicit a narrative discourse in the form of a picture description. Written instructions were provided in both the native language and English. Each speaker was allowed sufficient time to examine the picture before elicitation. The same picture was used in both spoken languages for each bilingual.

Monologue

To elicit participants' spontaneous connected speech, all speakers were instructed to talk briefly about a selected topic, either an 'introduction to their family', or their 'life experience in United States'. The speech sample was obtained for a period of approximately one to two minutes. The same topic was used for speech samples in both the native language and English for each participant.

Recordings

All speech samples in the experiment were recorded on a digital audiotape (DAT) recorder (Tascam DR-40) and sampled in a sound-attenuated booth using a high quality head-mounted microphone (Sennheiser). The recordings were digitized at a sampling rate of 44. 1 kHz and 16 bits/sample quantization, and stored in a laptop for subsequent analyses. The microphone-to-mouth distance for each talker was maintained to be approximately 3 – 4 cm at all times, and slightly to the side of the speaker's mouth.

Accent rating in English

In addition to the selection of participants based on screening interviews, the degree of foreign accent in English was assessed for each speaker. The purpose of this task was to further determine oral language proficiency by evaluating the degree of detectable non-native accent in second language production. A listening test was devised in order to obtain measures of degree of accent in the participating speakers.

The rating task was performed by 20 native speakers of English (three males, 17 females). Their ages ranged from 21 to 51 years (a mean of 31.9 years, SD = 4.35). All listeners were born and educated in the United States and ranged in education from 12 to 18 years

Table 1 Demographic information on participants in Korean-English (KE) group and Mandarin-English (ME) group; mean values and standard deviation (in parentheses) are provided

Group	Total (*n*)	Age	General education (years)	English learning (years)	Age of arrival (AoA) in the U.S.	Years of residence in the U.S.	Self-rated English oral fluency (1 – very poor, 5 – near-native)	Self-rated overall English proficiency (1 – very poor, 5 – near-native)
KE	11	25.5 (1.75)	17.18 (1.25)	13.55 (2.83)	12.91 (7.25)	10.59 (4.33)	4.18 (0.87)	4.36 (0.67)
ME	11	23.4 (0.81)	17.27 (0.47)	13.32 (3.19)	22.18 (0.98)	1.22 (0.41)	3.55 (0.69)	3.73 (0.65)

(SD = 1.89). Participants with or without a background in communicative sciences were included. Participants were excluded if they had prior language knowledge on Mandarin or Korean. Each listener participated in the experiment privately in a quiet room using a software program EXPERIGEN (Becker and Levine, 2014) that enabled implementation of the study and data collection via computer interface. All listeners were informed that a series of speech samples would be played to them one at a time and their task was to assess how strong they perceived the foreign accent to be. Based on a 9-point Likert scale (1 – very strong foreign accent, 9 – no foreign accent), a total of 132 speech stimuli was assessed by listeners. After hearing an utterance, a visual scale from 1 to 9 appeared on the computer screen and each participant clicked on a number that matched to their decision of degree of accent. Before the testing started, two practice trials were provided with speech samples (which were excluded from the experiment) produced by native speakers of English.

The stimuli in the experiment were 6–10 second speech recordings excerpted from the original experiment, all representing the Korean-English and Mandarin-English speakers' speech in English. Three utterances were selected each from the reading task and spontaneous speech (a total of six utterances per speaker), respectively. This was intended to examine whether there would be a difference between the speech tasks in perception of accent and to provide additional information on the bilingual speakers. It has been reported that reading is rated as more strongly accented than spontaneous speech (Oyama 1976; Thompson, 1991) and we intended to examine whether the task effect is maintained in our accent ratings. The order of presentation of the trials was randomized for each listener, and the stimuli could be replayed if needed.

Data analysis

Acoustic measures

Acoustic vocal measures, including average F0, F0 variability (coefficient of variation; standard deviation divided by mean), average intensity (dB SPL), and speech rate (syllables per sec) were obtained from the speech sample recordings. Regarding speech rate, since factors such as dysfluencies and pause-fillers are frequently detected in natural speech (Cucchiarini *et al.*, 2002), values were obtained only for the reading task. Measures were acoustically analyzed from the recorded speech samples utilizing *Praat* (Boersma, 2001), a speech analysis software program, using a pitch tracking system with a standard range setting of 75– 500 Hz. For acoustic data analysis, participants' responses were segmented into multiple phrasal or sentential utterances. Silence periods between phrases and sentences were edited if they were beyond one second. Values of the measures for all segmented portions of connected speech were obtained and averaged. All measures were manually checked and corrected for artifacts (e.g. spurious F0 values outside the range of normal pitch).

Statistical analyses

For the accent rating task, a two-way mixed analysis of variance (ANOVA), using SPSS software program for statistical analyses, was conducted to compare the degree of foreign accent between KE and ME group in the reading task and monologue. Additionally, for group comparison, independent *t*-tests and chi-square tests were conducted on the demographic information of the KE and ME groups. Regarding the acoustic measures, a series of 3-way mixed ANOVA and 2-way mixed ANOVA tests were conducted on the data. The measures of F0, F0 variability, and intensity were analyzed using three-way mixed ANOVA with Language (native language vs. English) and Task (reading, picture description, monologue) as within-group variables, and Group (KE group vs. ME group) as a between-group variable. Speech rate was analyzed using a two-way mixed ANOVA with Language (native language vs. English) and Group (KE group vs. ME group) as factors. We considered an alpha level of 0.05 to be statistically significant in all following analyses. The Bonferroni's test was used to carry out post hoc pairwise comparisons, when the ANOVA results indicated significant effects within or between variables.

Results

Demographic comparisons

Comparisons were made between the KE and ME groups based on the data obtained in the written

questionnaire (see Table 1). The following factors of interest were analyzed: age, years of general education, years of English education, AoA in the States, years of residence in the States, self-rated English oral proficiency (how fluent one speaks in English), self-rated overall English competency (overall command of English). Self-ratings on the participants' English competency and oral proficiency were made on a scale from 1 (very poor) to 5 (near-native).

Results revealed that KE group and ME group did not differ in number of years of general education and years of English learning. Also, for self-rated English oral proficiency, independent *t*-tests failed to reveal significant differences between the KE and ME groups ($P = 0.149$). For the scores on self-rated English overall proficiency, however, the results yielded a significant difference between the KE and ME groups ($P = 0.04$), as the KE group perceived their overall proficiency in English to be better than the ME group. The two groups showed significant differences in three additional categories: age ($P < 0.001$), AoA in the US ($P < 0.005$), and years of residence in the US ($P < 0.005$). Regarding age, the mean age of the ME group was slightly lower than the mean age of the KE group (gap of 2.1 years), but the participants' ages ranged between 21 and 27 years, which are not expected to display distinct vocal differences due to physiological changes based on aging effect (Boulet and Oddens, 1988; Sataloff and Linvelle, 2006). The average AoA of KE group was lower than ME group, indicating that the Korean-English bilinguals were exposed to a richer English environment, arriving in the U.S. at a mean age of 12.91 years, whereas the participants in ME group arrived in the U.S. at a mean age of 22.18 years. A corresponding difference occurred in the years of residence in the U.S. also. Specifically, KE group had an average of 10.59 years (SD = 7.25) for the years of residence, which was significantly higher than the mean of ME group of 1.22 years (SD = 0.41).

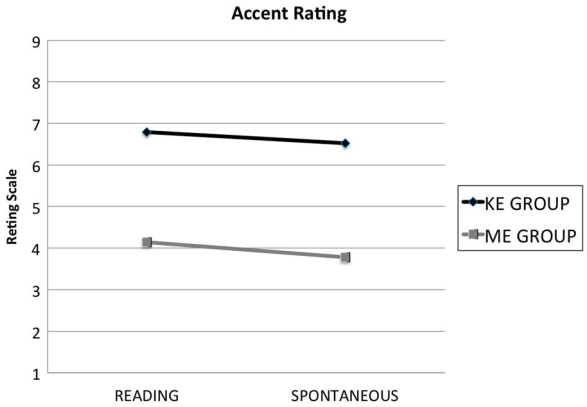

Figure 1 Ratings of the degree of accent in English between KE and ME group across two speech tasks are shown in line graphs, respectively. Results show significant difference between KE and ME group across both tasks, and between reading and spontaneous speech for KE group only ($P < 0.05$).

Overall, group comparison results reveal that both groups were fairly competent in oral production of their second language, considering English education years and self-ratings of English oral fluency. However, differences were noticed between the two bilingual groups such as AoA, years of residence in the U.S., and self-ratings of English proficiency. Along with the results for the accent ratings, this issue will be further addressed in the discussion section.

Accent rating in English

The foreign accentedness in the English speech samples of KE and ME group were analyzed using a two-way mixed ANOVA with Task (reading vs. spontaneous) and Group (KE vs. ME group) as factors. The mixed ANOVA results yielded significant main effects of Task (F (2,20) = 7.297, $P < 0.005$) and Group (F (1,20) = 1468.518, $P < 0.005$), but no significant interaction effect of Task × Group (F (2,20) = 0.165, $P = 0.847$). Fig. 1 depicts the average rating values across tasks and groups. As shown, KE group had a significantly higher rating (less accented speech) compared to the ME group in both tasks. This indicates that, despite both groups showing high level of English on other measures, the accent rating task provided further information on second language proficiency in the oral production, with the KE group perceived as having less foreign accent than the ME group. For the KE group, a significant task difference between accent ratings was also found; accent was rated on an average of 4.37 (SD = 0.27) for reading and 3.93 (SD = 0.40) for spontaneous speech ($t(20) = 2.007$, $P = 0.06$), indicating that the speech samples from the reading task was perceived as being less accented than spontaneous speech. For the ME group, however, the average rating of their English was 6.78 (SD = 0.15) in reading, whereas they were rated at a mean of 6.53 (SD = 0.54) for spontaneous speech, with no significant difference ($t(20) = 2.246$, $P = 0.54$).

The two groups differed in their native languages, and therefore the effects of language characteristics on perceived pronunciation cannot be ruled out. However, the rating differences were likely influenced at least in part by demographic factors. It can be assumed from the demographic data, for whom some pertinent characteristics are equated, that AOA and that years of residence in the U.S. may have influenced the results of the rating protocol (Flege *et al.*, 1995).

Mean Fo

A three-way mixed ANOVA revealed a significant Language effect, $F(1, 20) = 14.79$, $P = 0.001$ on F0

measures, indicating an overall significantly higher average F0 in the native language compared to English (Korean, $M = 213.61$, SD = 3.21; Mandarin, $M = 205.38$, SD = 2.91) in both KE and ME groups. A significant main effect of Task was also found ($F(2, 20) = 37.4$, $P < 0.0005$), with the reading task having the highest mean F0 values, followed by the picture description and the monologue task. *Post hoc* pairwise comparisons revealed that there was a significant difference between reading and picture description ($P < 0.0005$), and reading and monologue ($P < 0.0005$), but picture description and monologue did not show significant differences on F0 values. There was no main effect of Group or interaction between Group and Language, indicating that there was no overall group difference on the average F0 values ($F(1,20) = 0.67$, $P = 0.421$) nor was there a group difference between the F0s of native language and English ($F(1,20) = 1.792$, $P = 0.196$). However, the results yielded significant interaction effects of Language × Task, $F(2,20) = 5.357$, $P = 0.009$), and Group × Task, $F(2,20) = 8.624$, $P = 0.001$, reflecting that, depending on the given task, there was a difference between native language and English, and the groups differed in their performance between the two spoken languages; *post hoc* tests showed that the KE group consistently exhibited a significantly higher mean F0 in Korean compared with English in all three tasks; for the ME group, however, the average F0 was significantly higher in Mandarin than in English only in the reading task ($P = 0.014$), indicating that differences in F0 between speaking Mandarin and English depended on the type of speech task. The effects of F0 values associated with spoken languages and tasks in the KE and ME groups are depicted in Fig. 2.

Fo variability

To measure F0 variability, coefficient of variation (standard deviation divided by mean) was used to estimate the relative magnitude of the variance from the mean F0. Results indicated a significant main effect of Language ($F(1,20) = 6.073$, $P = 0.023$). *Post hoc* pairwise comparisons showed that the KE group exhibited a significantly greater pitch variability in Korean compared with English ($P = 0.01$), as depicted in Fig. 3. For the ME group, F0 variability was similar between Mandarin and English for all three tasks. There were no significant effects of Group or Task, or any significant interaction effects on F0 variability, indicating that task selection did not affect the variability in F0 for both groups.

Intensity

ANOVA results revealed a significant main effect of Group ($F(1,20) = 11.84$, $P < 0.003$) on intensity. Fig. 4 illustrates the average intensity (dB) values for the two spoken languages across all tasks for both groups. However, neither Language nor Task, or their interaction, had any impact on intensity. The ME group showed a higher average intensity compared to the KE group with a mean difference value of 4.99 dB, indicating that the ME group were measured as significantly louder, but the intensity between native language and English was comparable across tasks and languages.

Speech rate

The 2 × 2 mixed ANOVA results yielded significant main effects of Language ($F(1,20) = 48.678$, $P < 0.0005$) and Group ($F(1,20) = 12.283$, $P < 0.0005$), and a significant interaction effect of Language × Group ($F(1,20) = 12.283$, $P = 0.002$). Fig. 5 shows the average rate (syllables per second) across native language and English. Participants spoke significantly faster in their native language compared to English. For KE group, Korean was spoken at a rate of 5.18 syllables per second (SD = 0.45), whereas English was spoken at a mean of 4.03 syllables per second (SD = 0.54). For ME group, average rate for Mandarin was 4.01 syllables per second (SD = 0.27) and 3.63 syllables per second (SD = 0.40) for English. Results showed that speech rates between KE and ME group also differed significantly. The rate difference between the two languages was greater in the KE group than in the ME group. When rates for English were compared using *t*-tests, however, there was only a marginally significant difference ($t(20) = 2.007$, $P = 0.06$), indicating that the speech rate for English was comparable between groups.

Discussion

This study examined the voice profiles of bilingual speakers when speaking two different languages across speech tasks in order to address the possible difference in vocal parameters related to language selection and type of speech task. F0, F0 variability, intensity, and rate, parameters of importance in judging a voice pattern, were the measures selected for this study. Findings indicated that there was a significant effect of language and task on these selected vocal characteristics in bilingual speakers. With respect to language effect, the bilingual speakers showed significant differences in the average F0, F0 variability, and speech rate in their two languages. Furthermore, there was an interaction effect between language selection and task on F0, suggesting that the difference in F0 between languages was based on the speech task that the bilinguals performed. However, intensity did not reveal any language or task effect, but showed the ME group speaking overall significantly louder than the KE group.

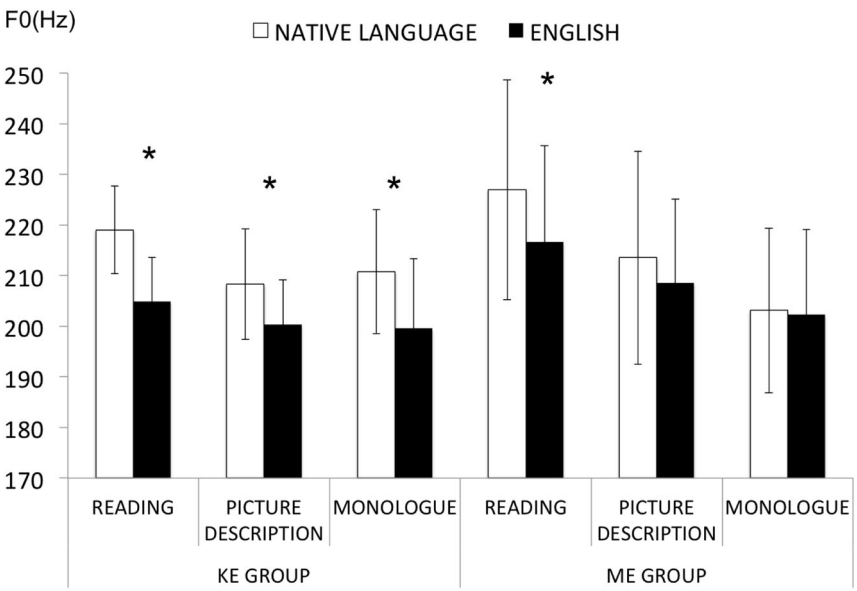

Figure 2 Average F0 (Hz) and F0 standard deviations (error bars) between two spoken languages across three speech tasks are shown in bar graphs for KE (left) and ME (right) group, respectively. Statistical significance is indicated by an asterisk ($P < 0.05$).

Speech rate differed between two spoken languages within both groups, but also between the groups. Overall, the findings provide further support that the bilingual voice features consistently different vocal patterns, contingent not only on the language (native versus acquired) that is produced, but also on the type of speech task. We turn to provide some explanations of the outcomes and for the resulting profile of a bilingual voice.

Concerning F0 measures, a significantly higher F0 in either Korean or Mandarin than in English was evident – concurring with findings reported in some previous studies (Keating and Kuo, 2012; Mang, 2001). This contrasts with other bilingual studies that found a higher F0 in English than in the native languages (Jarvinen *et al.*, 2013; Ng *et al.*, 2010; Ng *et al.*, 2012), suggesting that using a foreign language can lead to possible higher F0 due to emotional factors (Jarvinen *et al.*, 2013). It is speculated that all participants in the study were overall fluent and sufficiently competent in oral English so that the effect of speaking a foreign language did not strongly affect a change in the F0 (i.e. increased F0). In that regard, the average F0 values in English between the two groups were not significantly different, suggesting that the difference in F0s based on language selection cannot be fully explained by simply the influence of using a second language.

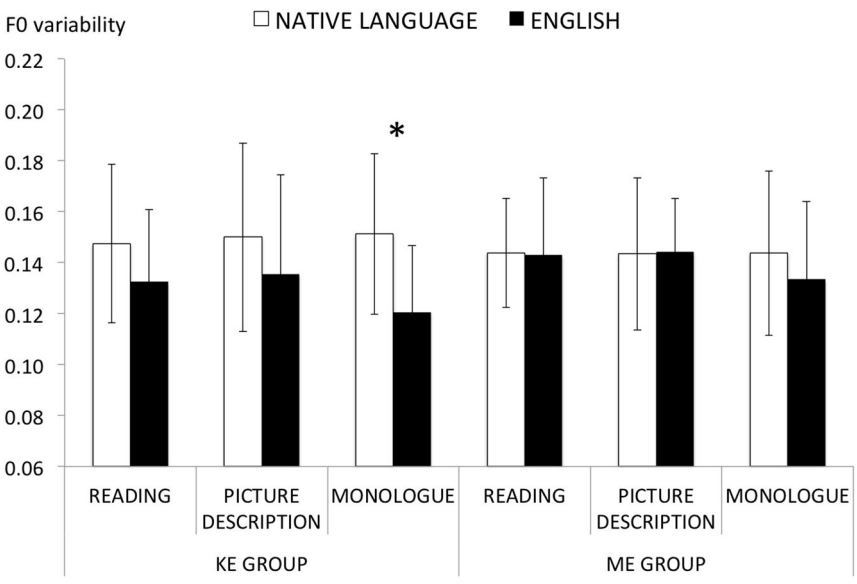

Figure 3 Utterance-level mean F0 variability (coefficient of variation) between two spoken languages across speech tasks is shown in bar graphs for KE (left) and ME (right) group, respectively. Statistical significance is indicated by an asterisk ($P < 0.05$).

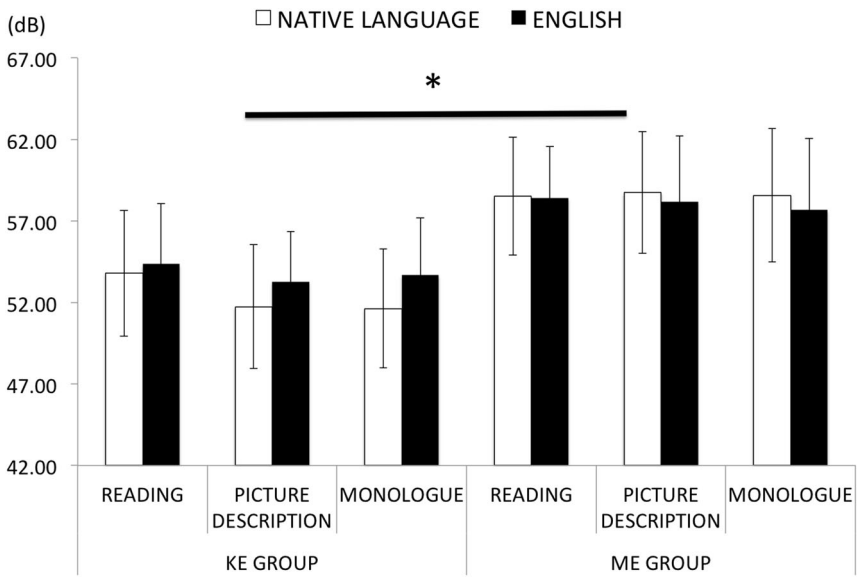

Figure 4 Utterance-level mean intensity (dB) between native language and English across three speech tasks is shown in bar graphs for KE (left) and ME (right) group, respectively. Statistical significance is indicated by an asterisk ($P < 0.05$).

Rather, other factors should be of consideration. If a bilingual speaker is orally proficient in two languages, it may be that the intrinsic acoustic features of each language contribute to the difference in the acoustic vocal parameters. Specifically, acoustic phonetic characteristics of each language may have influenced vocal changes in overall speech. As mentioned, a higher F0 for tonal languages was demonstrated, compared with non-tonal languages, not only during isolated vowel productions (Eady, 1982; Mang, 2001) but also in connected speech (Keating and Kuo, 2012; Ng *et al.*, 2012). The current findings support the notion of a normative higher F0 in tonal languages (Mandarin) than in non-tonal ones (English). In the case of the Korean language, also, the mean F0 in Korean was significantly higher than in English. Although there are no reported cross-linguistic studies that compare the acoustic pitch features of Korean and English in connected speech, we consider the consonant-tone interaction and tense consonants in Korean to be a possible cause of the similar phonetic effects of higher F0, in a manner parallel to that of tonal languages. Considering that tense consonants in Korean but not in English bring forth a higher F0, the difference in F0 between Korean and English could be attributed to the intrinsic F0 features of tense and tonal effect of Korean consonants. The Korean-English bilingual speakers may adequately

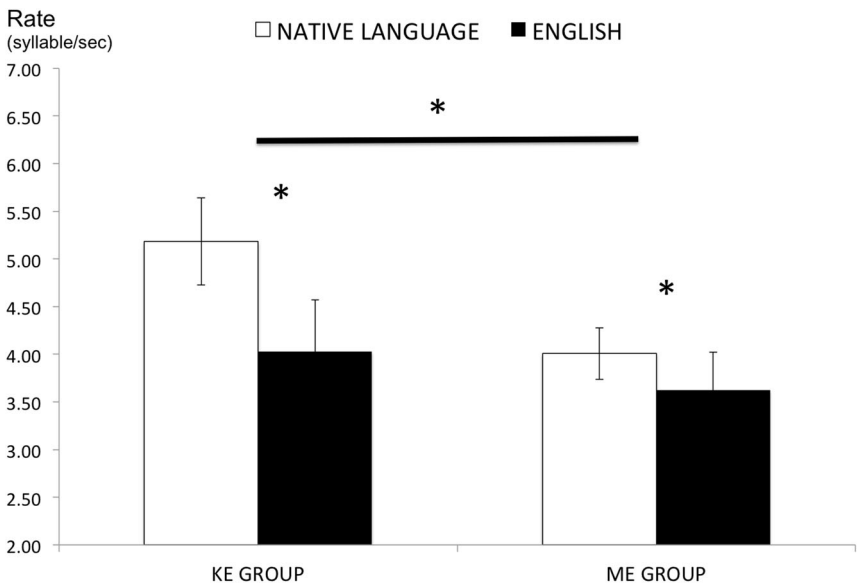

Figure 5 Utterance-level mean speech rate (syllables per second) and standard deviations (error bars) of speaking rate between native language and English in reading task are shown in graphs for KE (left) and ME (right) group, respectively. Statistical significance is indicated by an asterisk ($P < 0.05$)*.

produce the tense consonants and tonal phonetic effects of consonants in Korean during connected speech, resulting in overall higher F0. We speculate that the vocal differences between the two spoken languages in the KE group may derive from possible linguistic factors such as this difference in phonetic repertoire of consonants. Therefore, both Korean and Mandarin resulted in overall higher F0 than English. Further studies that explore the relationship between tone and tense in languages and overall pitch change in the natural speech of bilingual speakers should be considered.

Interestingly, it is noted that whereas the KE group showed significant differences between their two spoken languages across all speech tasks, the ME group demonstrated significant difference only during the reading task. There can be two explanations for the different F0 outcomes across groups. First, reading is a structured task that proven to be a robust measure for eliciting comparatively higher F0 values (Hudson and Holbrook, 1982; Mysak, 1959; Ramig and Ringel, 1983; Sorenson, 1989). It may be that the nature of performing a reading task can underscore the acoustic attributes of each language since the speakers may hyper-articulate with better control of speech, resulting in a stark F0 contrast between the two spoken languages. Another reason may lie in the difference in oral English proficiency observed in the accent ratings. Although overall English skills were competent among the two groups, a stronger foreign accent in English for the ME group may have led to comparable F0s between Mandarin and English on their spontaneous speech provide partial explanation to the F0 difference in two languages across tasks. Various aspects of oral proficiency in a language may certainly have an impact on the acoustic patterns in the bilingual voice. The effect of second language oral proficiency on voice in many of its perspectives remains to be explored further.

With regard to speech rate, there was a significant difference between the two spoken languages for both groups. This can be attributed to similar issues dealt in the results of either the intrinsic linguistic features of each language or F0 relating to second language oral proficiency. Regarding the first reason, a recent study conducted by Pellegrino and her colleagues (2011) shed light on how the speed of spoken production can be related to syllabic rate and information density. According to their study, there exists a negative correlation between speaking rate based on syllabic structure and syllabic information density. Languages that have a high speaking rate have a tendency to pack less semantic information into each syllable (decreased information density) and vice versa. For instance, Spanish has a high speech rate with low density syllables, whereas Mandarin is considered a language with slower syllabic rate with high information density (due to tonal features packed in syllables). The present study is in agreement with Pellegrino *et al.* (2011) in that Korean exhibits higher syllabic speaking rates compared with English and Mandarin, which are characterized by slow syllabic rate with high density syllables. Although the information rate theory does not provide a comprehensive explanation of the overall acoustic differences, it suggests that language-specific syllable density may influence measures of speaking rate. Considering the speech rate findings to be related to second language use, it is mentioned that overall second language proficiency can influence the rate in natural speech (Munro and Derwing, 2001). That is, a slower rate in the second language, compared to the native tongue, is the result of the higher demands of the cognitive task and a lack of competency in oral control. We confirmed in this study that both groups demonstrated a slower rate in English, which may be attributed to the fact that all of the speakers are non-native speakers of English. The F0 measure may not have been affected by the second language proficiency, but it may have for the speech rate. Further investigation is warranted to determine how linguistic factors influence speaking rates across languages.

Another possible explanation for the different voice profiles in bilinguals results from a consideration of socio-cultural or pragmatic factors that can influence vocal features within a language (van Bezooijen, 1995). Changing vocal quality within individuals can occur with various pragmatic functions (Kreiman and Sidtis, 2011; Zraick *et al.*, 2005). Loveday (1981), for instance, explored F0 differences in Japanese and English speakers in both males and females, and found that Japanese females adopted a high F0 that is very distinct from both English speakers and Japanese males. This was attributed to socio-cultural expectations and the pragmatic purpose of expressing politeness or femininity. Studies on the phonetic profile of politeness revealed that Korean native speakers also show a marked distinction in average F0 between different levels of polite speech, but in this case greater politeness was correlated with lower average F0 (Brown *et al.*, 2014; Shin, 2005; Winter and Grawunder, 2012). Nonetheless, the fact that Korean female speakers share with Japanese females the cultural values of exhibiting expected politeness and femininity within society (Han, 1992) indicates that socio-cultural factors may influence changes in vocal quality. The potential correlation between cultural characteristics in certain language groups and acoustic vocal features of speech seems to merit further investigation.

In respect to task effect on vocal features, early studies on voice have demonstrated that monolingual speakers have a higher F0 associated with a structured task compared with an unstructured one (Hollien *et al*., 1997; Hollien and Jackson, 1973). Our findings show that this effect of task applies consistently to bilinguals. This observation was also reflected in the accent ratings, where reading was rated as significantly less accented than spontaneous speech. Task effect may be evident in both perception and production of speech. The inconsistent findings from the previous literature may have resulted from the restricted or selective methodological approach of using a specific task without comparison to other tasks.

The present study is distinct from previous works by its inclusion of a number of vocal parameters and the attention to speech task type in establishing a more refined bilingual voice profile. However, caution should be taken not to generalize the findings to the entire bilingual population, since only a limited number of females were included in the study; there may well be differences between genders in use of voice on tasks. Furthermore, as mentioned above, although language competency in the two spoken languages was overall well-balanced, subtle differences in oral English pronunciation skills occurred between groups. Based on the language profiles of the ME group, this is somewhat predictable as there is a positive correlation between detectable foreign accents and the age of learning in a second language (Flege and Fletcher, 1992; Thompson, 1991). Accented production in a language can induce differences in phoneme sets and intonation patterns, and create overall changes in pronunciation patterns (Arslan and Hansen, 1996). It is not known to what extent degree of foreign accent influences vocal characteristics in bilinguals and this merits exploration in future studies. In order to do so, comparison of groups of monolinguals, and bilinguals with same and different native language, should be explored in details. Despite caveats, the findings here provide support for the growing body of literature showing that the use of different languages clearly manifests different outcomes of vocal characteristics, even within bilingual speakers.

Conclusion

The present study attempts to explore vocal differences in bilingual speakers that occur, depending on their choice of language and type of speech task. Findings revealed that bilingual speakers produced notably different voice patterns in their two spoken languages across different speech tasks, when vocal measures were explored. The difference implies that language alone is an acquired factor that contributes to the manifestation of within- and between-speaker variability of vocal attributes. In addition, results indicate that selection of speech sample can be a crucial factor in differentiating acoustic profiles, supporting previous literature that task effects emerge in acoustic measures within individuals (Zraick *et al*., 2005). Future studies will do well to further investigate vocal features of bilingual speech relating to factors such as gender, age, and language dominance, for a better understanding of bilingualism and its effect on natural speech.

Disclaimer statements

Contributors None.

Funding None.

Conflicts of interest None.

Ethical approval None.

ORCID

Binna Lee http://orcid.org/0000-0001-5320-9703

References

Altenberg, E.P., Ferrand, C.T. 2006. Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice*, 20: 89–96.

Amino, K., Osanai, T. 2015. Cross-language differences of articulation rate and its transfer into Japanese as a second language. *Forensic Science International*, 249: 116–122.

Andrianopoulos, M.V., Darrow, N.K., Chen, J. 2001. Multimodal standardization of voice among four multicultural populations: fundamental frequency and spectral characteristics. *Journal of Voice*, 15: 259–272.

Arslan, L., Hansen, J. 1996. A study of temporal features and frequency characteristics in American English foreign accent. *The Journal of the Acoustical Society of America*, 102(1): 28–40.

Asher, J.J., García, R. 1969. The optimal age to learn a foreign language. *The Modern Language Journal*, 53(5): 334–341.

Awan, S.N., Mueller, P.B. 1996. Speaking fundamental frequency characteristics of white, American, and Hispanic kindergarteners. *Journal of Speech, Hearing, and Research*, 39: 573–577.

Baken, R., Orlikoff, R. 2000. *Clinical measurement of speech and voice*. San Diego, CA: Singular Thomson Learning.

Baker, S., Weinrich, B., Bevington, M., Schroth, K., Schroeder, E. 2008. The effect of task type on fundamental frequency in children. *International Journal of Pediatric Otorhinolaryngology*, 72 (6): 885–889.

Becker, M., Jonathan, L. 2014. Experigen – an online experiment platform. Available at http://becker.phonologist.org/experigen.

Bialystok, E. 1997. The structure of age: in search of barriers to second language acquisition. *Second Language Research*, 13 (2): 116–137.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5: 341–345.

Boulet, M.J., Oddens, B.J. 1988. Female voice changes around and after the menopause – an initial investigation. *Maturitas*, 23(1): 15–21.

Brown, L., Winter, B., Idemaru, K., Grawunder, S. 2014. Phonetics and politeness: perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, 66: 45–60.

Cucchiarini, C., Strik, H., Boves, L. 2002. Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *The Journal of the Acoustic Society in America*, 111(6): 2862–2873.

Eady, S. 1982. Differences in the F0 patterns of speech: tone language versus stress language. *Language and Speech*, 25: 29–42.

Fairbanks, G. 1960. *Voice and articulation drillbook*. New York: Harper and Row.

Flege, J., Fletcher, K. 1992. Talker and listener effects on the perception of degree of foreign accent. *The Journal of the Acoustic Society in America*, 91: 370–389.
Flege, J., Munro, M., Mackay, I. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustic Society in America*, 97: 3125–3134.
Gelfer, M., Denor, S. 2014. Speaking fundamental frequency and individual variability in Caucasian and African-American school-aged children. *American Journal of Speech-Language Pathology*, 23(3): 395–406.
Goodglass, H., Kaplan, E. 1972. *The assessment of aphasia and related disorders*. Lea and Febiger: Philadelphia.
Han, C. 1992. A comparative study of compliment responses: Korean females in Korean interactions and in English interactions. *Working Papers in Educational Linguistics*, 8: 17–31.
Hanley, T.D., Snidecor, J.C., Ringel, R.L. 1966. Some acoustic differences among languages. *Phonetica*, 14: 97–107.
Hollien, H., Jackson, B. 1973. Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics*, 1: 117–20.
Hollien, H., Hollien, P.A., Jong, G. 1997. Effect of three parameters on speaking fundamental frequency. *The Journal of the Acoustical Society of America*, 102(5): 2984–2992.
Hudson, A.I., Holbrook, A. 1982. Fundamental frequency characteristics of young black adults: spontaneous speaking and oral reading. *Journal of Speech and Hearing Research*, 25: 25–28.
Hunter, E. 2009. A comparison of a child's fundamental frequencies in structured elicited vocalizations versus unstructured natural vocalizations: a case study. *International Journal of Pediatric Otorhinoloaryngology*, 73: 561–571.
Jarvinen, K., Laukkanen, A., Aaltonen, O. 2013. Speaking a foreign language and its effect on F0. *Logopedics Phoniatrics Vocology*, 38: 47–51.
Jun, S. 1993. *The phonetics and phonology of Korean prosody* (Doctoral dissertation, The Ohio State University).
Keating, P., Kuo, P. 2012. Comparison of speaking fundamental frequency in English and Mandarin. *Journal of Acoustic Society of America*, 132: 1050–1060.
Kreiman, J., Sidtis, D. 2011. *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Boston: Wiley-Blackwell.
Loveday, L. 1981. Pitch, politeness and sexual role: an exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24: 71–89.
MacKay, I.R., Meador, D., Flege, J.E. 2000. The identification of English consonants by native speakers of Italian. *Phonetica*, 58(1–2): 103–125.
Mang, E. 2001. A cross-language comparison of preschool children's vocal fundamental frequency in speech and song production. *Research Studies in Music Education*, 16: 4–14.
Mayo, R., Manning, W.H. 1994. Vocal tract characteristics of African-American and European-American adult males. *Texas Journal of Audiology and Speech-Language Pathology*, 20: 33–36.
McNamara, T.F. 1996. *Measuring second language performance*. London: Addison Wesley Longman.
Mennen, I., Schaeffler, F., Docherty, G. 2012. Cross-language differences in fundamental frequency range: a comparison of English and German. *The Journal of the Acoustic Society of America*, 131: 2249–2260.
Morris, R.J. 1997. Speaking fundamental frequency characteristics of 8- through 10-year-old white- and African-American boys. *Journal of Communication Disorders*, 30: 101–116.
Munro, M.J., Derwing, T.M. 2001. Modeling perceptions of accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition*, 23: 451–468.
Mysak, E.D. 1959. Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research*, 2: 46–54.
Ng, M., Chen, Y., Chan, Y.K. 2012. Differences in vocal characteristics between Cantonese and English produced by Cantonese-English bilingual speakers – a long-term spectral analysis. *Journal of Voice*, 26: e171–e176.
Ng, M., Hsueh, G., Leung, C. 2010. Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International journal of speech-language pathology*, 12(3): 230–236.
Ohala, J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, 41: 1–16.
Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5(3): 261–283.
Pellegrino, F., Coupe, C., Marsico, E. 2011. Across-language perspective on speech information rate. *Language*, 87(3): 539–558.
Ramig, L.A., Ringel, R.L. 1983. Effect of physiological aging on selected acoustic characteristics of voice. *Journal of Speech and Hearing Research*, 26(1): 22–30.
Reich, A.R., Mason, J.A., Frederickson, R.R., Schlauch, R.S. 1989. Factors influencing fundamental frequency range estimates in children. *Journal of Speech and Hearing Disorders*, 54(3): 429–438.
Roberts, P.M., Shenker, R.C. 2007. Assessment and treatment of stuttering in bilingual speakers. In: Conture E, Curlee R.F. (eds.) *Stuttering and related disorders of fluency* (3rd ed.) New York: Thieme Medical Publishers, Inc., pp. 183–209.
Sapienza, C.M. 1997. Aerodynamic and acoustic characteristics of the adult African American voice. *Journal of Voice*, 11: 410–416.
Sataloff, R.T., Linvelle, S.E. 2006. *Vocal health and pedagogy: advanced assessment and treatment*. 2nd ed. San Diego: Plural.
Shin, S. 2005. *Grammaticalization of politeness: a contrastive study of German, English and Korean*. Berkeley: Doctoral thesis, University of California.
Sorenson, D.N. 1989. A fundamental frequency investigation of children ages 6–10 years old. *Journal of Communication Disorders*, 22(2): 115–123.
Thompson, I. 1991. Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning*, 41: 177–204.
Todaka, Y. 1993. *A cross-language study of voice quality: bilingual Japanese and American speakers*. Los Angeles: Doctoral dissertation, University of California.
van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3): 253–256.
Wheat, M.C., Hudson, A.I. 1988. Spontaneous speaking fundamental frequency of 6-year-old black children. *Journal of Speech, Language, and Hearing Research*, 31(4): 723–725.
Winter, B., Grawunder, S. 2012. The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, 40(6): 808–815.
Xue, S.A., Needley, R., Hagstrom, F., Hao, J. 2001. Speaking F0 characteristics of elderly Euro-American and African-American speakers: building a clinical comparative platform. *Clinical Linguistic Phonetics*, 15: 245–252.
Yamazawa, H., Hollien, H. 1992. Speaking fundamental frequency patterns of Japanese women. *Phonetica*, 49: 128–140.
Zraick, R.I., Birdwell, K., Smith-Olinde, L. 2005. The effect of speaking sample duration on determination of habitual pitch. *Journal of Voice*, 19: 197–201.