Vocal characteristics across English-Northern Sotho bilingual speakers: a comparative study

Amy Hammann[a], Bhavani Pillay[b], Marien Alet Graham [c], Jeannie van der Linde[d]

[a] Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

[b] Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

[c] Department of Science, Mathematics and Technology Education, University of Pretoria, Pretoria, South Africa

[d] Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

Short Title: The English-Northern Sotho Bilingual Voice

Corresponding Author: Jeannie van der Linde

Department of Speech-Language Pathology and Audiology

University of Pretoria

Pretoria, Gauteng, 0028, South Africa

Tel: 012 420 2948

E-mail: jeannie.vanderlinde@up.ac.za

Number of Tables: 4

Number of Figures: 0

Word count: 3, 147

Keywords: voice, bilingualism, English, Northern Sotho, cross-linguistic effect

Abstract

Introduction: Bilinguals constitute a significant portion of speech-language pathologists' (SLPs) caseloads. Insight into the cross-linguistic effect on voice is needed to guide SLPs to make linguistically-appropriate observations when working with heterogenous populations.

Method: Nineteen female English-Northern Sotho bilinguals performed three speech tasks (reading, picture description and monologue) in each language. Acoustic analysis of mean fundamental frequency (f_0), intensity, and articulation rate was conducted with Praat. A panel of blinded listeners reached consensus after independently reviewing the recordings during perceptual analysis of voice quality, resonance, and glottal attack.

Results: The following statistically significant differences were found across and within the languages: The mean f_0 was 204.61Hz in the Northern Sotho picture description yet 196.50Hz in the English picture description. The mean intensity of reading in Northern Sotho was 66.38dB whereas the mean intensity of reading in English was 65.09dB. Articulation rate was 3.78 syllables/sec in English passage reading and 3.41 syllables/sec in Northern Sotho passage reading. Within English, passage reading elicited a significantly quicker articulation rate than the picture description (3.34 syllables/sec) and monologue (3.46syllables/sec). Within Northern Sotho, mean f_0 was 203.83Hz in passage reading yet 191.11Hz in the monologue. Perceptual voice quality, glottal attack, and resonance were comparable across languages.

Discussion/Conclusion:. Relationships between languages spoken, task performance, and vocal characteristics were observed in bilingual speakers. SLPs must consider the interaction thereof when working with bilingual voice clients.

Vocal characteristics across speech tasks in English-Northern Sotho bilingual individuals: a comparative study

Introduction

Speech-language pathologists (SLPs) require a holistic understanding of the complex nature of voice to best-serve the heterogenous population with whom they interact [1]. Ample research supports the connection between internal factors, such as age, gender, anatomy, emotional state, ethnicity and voice [2, 3, 5]. However, a small number of studies have explored the relationship between voice and external factors, such as language and speech tasks, when striving to understand voice in its entirety [4 – 6].

The extent to which language exclusively affects voice has received sporadic interest and remains unclear in recent research [7]. Interest in bilingualism continues to rise as bilingualism characterises contemporary society [9]. English has been the most studied language in cross-linguistic research; owing to its global dominance in education and business [8, 10]. Bilingual speakers' vocal characteristics have been explored across English and other languages including: Japanese, Hebrew, Welsh, Mandarin, Korean, Cantonese, Arabian, Finnish and Russian [4, 5, 13 – 16]. Fundamental frequency (f_0) has been the most widely considered parameter in cross-linguistic voice studies, and is a salient feature of talker identity [8, 16]. F_0 depends on the instantaneous configuration of the larynx as mediated by the language's typical vocal tendency [13, 21].

Significant f_0 differences were identified when equally proficient English-Japanese, English-Welsh, and English-Korean bilingual speakers read in their respective languages [5, 11, 11]. The English-Japanese bilinguals conversed in both languages indistinguishably from monolingual speakers of either language [5]. The English-Welsh speakers acquired both languages in childhood and routinely used both languages in adulthood [11]. The English-Korean bilingual speakers were deemed proficient by an interviewer who was proficient in both English and Korean [11]. Reading in Japanese evoked higher pitch levels and wider f_0 spans in male and female English-Japanese bilingual speakers [11]. The Welsh reading passage elicited wider f_0 span and higher maximum f_0 from female English-Welsh bilinguals compared to English [11]. Female English-Korean speakers presented with significantly higher f_0 when reading in Korean [5].

Fewer studies have recognised speech tasks as a confounding factor [4 – 7]. The structured nature of reading may cause speakers to hyperarticulate and present with higher f_0 measurements [5, 19, 33]. Spontaneous speech tasks, however, reflect the speaker's true vocal nature in conversation [4, 20]. Bilingual speakers may employ language-specific laryngeal configurations that induce f_0 variations in spontaneous speech [7]. Cross-linguistic f_0 variations in spontaneous speech suggest that the language spoken, rather than task-type, contributes to language-based vocal differences [4, 7, 13].

F_0 variability has been observed across language pairs of differing origin in spontaneous speech although the direction of change has been erratic. In spontaneous speech, female English-Hebrew bilinguals presented with significantly higher mean f_0 in Hebrew. Similarly, male English-Hebrew bilinguals presented with significantly lower mean f_0 in English spontaneous speech [4]. Female English-Cantonese bilingual speakers had significantly higher f_0 values in English

connected speech [13]. Researchers acknowledged that languages have different phonetic settings to accommodate their different phoneme inventories, phonation types, pitch ranges, register, and prosody. These language-specific configurations of the vocal apparatus are thought to contribute to changes in f0 across languages [12]. Additionally, there is convergence on the idea that intrinsic features of tonal languages precipitate higher mean f_0 [6, 11, 31]. Tonal languages present with increased lexical high tones in connected speech that could elevate f_0 measurements [31]. However, the inconsistent nature of differences in mean f0 across bilingual speakers suggest that f_0 variations alone do not provide sufficient insight into the language effect on voice.

In addition to differences in mean $f_{0,}$ listeners may perceive language-specific vocal differences across rate, intensity, and perceptual parameters [4, 18, 21]. English-Mandarin and English-Korean bilingual speakers yielded comparable speech rates in English but spoke significantly faster in Mandarin and Korean [6]. English spontaneous speech presented with greater incidence of hard glottal attack and vocal fry in comparison with Hebrew connected speech [4]. No task- or language-specific effect contributed to changes in intensity across English, Mandarin, and Korean [6].

Further multiparametric studies are needed to fully understand vocal characteristics as a function of language and speech tasks. A cross-linguistic voice study that investigated one of the African languages was not identified by the authors. Africa is a conducive context in which to investigate the language effect on voice as most of the population is thought to be multilingual [21]. It is estimated that 2, 110 languages are spoken across the continent [11, 22].

South African bilinguals present with an array of possible language pairs as 11 official languages are recognised [23]. Northern Sotho is one of the three most-spoken languages in South Africa [22]. Tonal languages such as Northern Sotho employ acoustic phonetic characteristics, including f_0 variation, aspiration, and devoicing, to index meaning through lexical tone[1]. Comparing the vocal characteristics of English-Northern Sotho bilinguals is expected to yield results comparable with the current body of research, while investigating a unique language pair. The following question is posed: do English-Northern Sotho bilingual speakers present with language-based vocal differences across speech tasks?

Materials & Method

A mixed between-within subject design was selected to quantify English and Northern Sotho vocal characteristics across a reading passage, picture description and monologue.

Participants

Nineteen English-Northern Sotho bilingual speakers volunteered to participate after viewing an advertisement about the study on social media platforms. The volunteers contributed to snowball sampling by suggesting prospective participants known to them. The inclusion criteria required participants to be: (1) female; (2) between 18 and 65 years old; (3) proficient in English and Northern Sotho; (4) presenting with normal hearing.

The hearScreen application provides clinically valid pure tone audiometric screening results in a timely manner [24]. The hearScreen application confirmed that each participant had normal hearing. Each participant completed a questionnaire that required them to disclose their age, and information about their voice and languages. The mean age of the participants was 22.5 (SD =

7.85) years old and their ages ranged from 18 years old to 54 years old. All participants felt that they had normal healthy voices and confirmed that they had never been diagnosed with a voice disorder. The participants were considered proficient bilinguals if they reported communicating regularly in both languages, with a variety of communication partners, across home, occupational, and social contexts [25].Participants estimated how old they were when they acquired each language. They indicated the environments and communication partners with whom they used each language on a typical day. It was seen as valuable to investigate these factors as language dominance may fluctuate across the lifespan based on language history, input and use [26]. This information is presented in Table 1. Including bilingual speakers allowed cross-linguistic comparisons to be made with each participant serving as their own control.

Procedure

Each participant was audio-recorded while performing a reading passage, picture description, and monologue in English and Northern Sotho. A Phillips VoiceTracer Audio Recorder (DVT1150) with standard settings was used to obtain audio recordings in a sound-attenuated booth. Verbal and written instructions were provided and required each participant to speak in their most natural comfortable voice. Participants were encouraged to take brief breaks between recordings to combat fatigue.

Speech tasks. The task-effect on voice is not novel [6]. The reading passage, picture description, and monologue tasks were selected to elicit semantically matched samples, in structured and unstructured tasks, across languages. The languages were counterbalanced, and the sequence of speech tasks was randomized to control for order effects. Some speech tasks, such as sustained vowel phonation and tone sweeps, rely solely on the physiology of voice so they do not

functionally represent voicing for speech [27, 28]. The authors of this study saw it valuable to investigate vocal characteristics in connected speech tasks that have been included in previous cross-linguistic studies [4, 6, 7].

The Reading Passage. The Rainbow Passage has been used in cross-linguistic studies investigating the language effect on voice [6, 27]. The original Rainbow Passage amounts to 330 words in 19 sentences in English [29]. It was translated to Northern Sotho and reviewed by a qualified linguist for this study. The Northern Sotho version amounted to 376 words in 19 sentences. Participants had the opportunity to practice reading each passage 2-3 times before being audio recorded.

The Picture Description. The "Cookie Theft" image from the Boston Diagnostic Aphasia Examination [30] has been used in previous cross-linguistic research as it elicits predictable abstract and concrete discourse from participants [6]. Video description and cartoon description tasks have been used in a similar manner [16, 17]. Participants acquainted themselves with the image and described the picture for approximately one minute in each language.

***The Monologue*.** A spontaneous, connected speech sample was elicited by asking participants to speak about a neutral topic, either "my family" or, "why I love South Africa" in both languages, for approximately 1 minute each. The same topic was used per participant across languages.

Data Analysis

The voice analysis protocol is presented in Table 2. The acoustic analysis was similar to Lee and Sidtis' [6] method. Praat software [32] was used to conduct acoustic analysis of mean f_0, intensity, and articulation rate across all recordings. Silent periods exceeding two seconds were

removed prior to analysis so that the articulation rate could be obtained as a sensitive estimate of actual speech execution time [34]. The perceptual analysis was similar to Nevo et al. [4]. Three SLPs comprised the blinded panel of listeners and conducted perceptual analysis. Each SLP initially made an independent judgment before reaching consensus. All the SLPs were familiar with English and Northern Sotho.The monologue was selected for perceptual analysis as it best represents natural communication [35].

Voice quality, or overall sound quality of the voice, was rated using the GRBASI 4-point rating scale [31]. The GRBASI scale rates the grade of overall voice quality, roughness, breathiness, asthenia, strain, and instability individually from 0, equalling normal, to 3, equalling severe. In the nominal scale, resonance was rated as nasal, oral, or throaty depending on the cavity in which most sound energy was perceptually produced [4]. Glottal attack refers to the initial adduction of the vocal folds and was characterized as adequate, soft, or hard [4].

Descriptive and inferential statistics were conducted using the Statistical Package for the Social Sciences. The Shapiro-Wilk test was used to test the continuous variables for normality. Since the underlying process distribution of the majority of the continuous variables differed significantly from normality (majority of Shapiro-Wilk p-values < 0.05), non-parametric tests were deemed appropriate when conducting statistics on the continuous variables. In recent statistics literature, there is a vast collection of non-parametric tests and these methods have been shown to perform well (or very similarly) to their normal theory counterparts [36, 37] and outperform their parametric counterparts under certain conditions. For example, researchers [36] have shown that nonparametric tests using ranking procedures outperform their parametric counterparts when the distribution is skewed or peaked [36], which is typically the case in practice. Further motivation for the use of non-parametric tests is the fact that $n < 30$. It is well-

known that the Central Limit Theorem (CLT) "comes to the rescue" when continuous variables are non-normal when the sample size is 30 or larger [38]; however, this is not the case for the current study. The median and interquartile range were considered along with the mean and standard deviation of mean f_0, intensity, and speech rate in the English and Northern Sotho speech tasks. The non-parametric Wilcoxon signed-rank (WSR) test was used to compare mean f_0, intensity, and speech rate as they presented across the English and Northern Sotho reading passages, monologues, and picture descriptions The non-parametric Friedman's test was used to determine whether there are significant differences between two or more related groups, for example, the Friedman's test was used to compare the acoustic parameters, of mean f_0, speech rate, and intensity, across speech tasks within the English group and conducted separately for the Northern Sotho group. Pairwise testing (represented in a standardized form by the letter Z) was conducted if significant task-based acoustic differences were found within either language in the Friedman's test. The GRBASI scale measured voice quality and the nominal scale rated resonance and glottal attack in the monologue tasks. The categorical perceptual variables obtained from the GRBASI and nominal scales were analyzed using McNemar's test (for 2x2 crosstabulations) and the test for marginal homogeneity (for crosstabulations greater than 2x2).

Results

The comparison of the voice characteristics in English-Northern Sotho bilingual speakers across speech tasks is presented according to the outcomes of the acoustic and perceptual analysis.

Acoustic analysis

Table 3 presents the statistical comparisons of mean f_0, intensity, and articulation rate across speech tasks and languages. The picture description task elicited a significantly higher mean f_0 in

Northern Sotho than in English (WSR = -2.938, p=0.002). Participants read with significantly greater intensity in Northern Sotho than in English (WSR = -2.113, p = 0.033). However, the participants read significantly faster in English than in Northern Sotho (WSR = -2.918, p = 0.002). No statistically significant differences were identified in the monologue task.

In English, the reading passage elicited a quicker articulation rate than the picture description task (Z = 3.488, p < 0.001) and the monologue (Z = 3.082, p = 0.002). In Northern Sotho, the participants presented with significantly higher mean f_0 during the reading task than in the monologue (Z = 3.407, p < 0.001). As shown in Table 4, pairwise comparisons for intensity were not conducted as intensity did not differ significantly across speech tasks within either language.

Perceptual analysis

The GRBASI scale and a nominal scale were used to conduct the perceptual analysis as shown in Table 2. No statistically significant language-based differences in voice quality, resonance, and glottal attack were identified between the English and Northern Sotho monologues.

Discussion/Conclusion

The study aimed to determine if acoustic and perceptual voice parameters differed across speech tasks in English-Northern Sotho bilingual females and between the two languages. Significant language- and task-effects across the acoustic parameters of mean $f_{0,}$ intensity, and articulation rate were identified. Perceptual voice quality, glottal attack, and resonance between the English and Northern Sotho monologues were comparable.

Northern Sotho speech had a higher mean $f_{0,}$ although the difference was insignificant when compared with English overall. Lee and Sidtis [6] and Cheng [33] found that English-Korean

female bilinguals presented with significantly higher mean f_0 during reading, picture description, and monologue tasks in Korean when compared with English. However, the mean f_0 was only significantly higher in the Northern Sotho picture description than in the English picture description. The picture description task required participants to use predictable language to construct original discourse. The fixed-but-flexible nature of the picture description task may have revealed f_0 differences across the languages that did not manifest during the unstructured monologue and the structured reading passage. These findings reiterate that task selection is an important factor to consider in cross-linguistic studies [8].

Intensity has seldom been compared across languages, although it communicates the speaker's emotional state and personality [6, 35]. Intonational accents, like those in Northern Sotho, specify f_0 and intensity targets [39]. This intrinsic feature of Northern Sotho was hypothesized to induce greater intensity and higher mean f_0 across all Northern Sotho speech tasks. However, only the reading passage ascribed to the anticipated trend. This suggests that the language-effect on voice cannot be solely delineated based on the intrinsic features of languages. No significant task effect on intensity was noted within the English samples and within the Northern Sotho samples. This finding corroborates with Lee & Sidtis [6] study; suggesting that the interaction between the language and speech task may moderate intensity to a greater extent than either language or task effects in isolation.

Significant mean f_0 differences were not identified between speech tasks in the English-language condition. However, significant differences in mean f_0 presented between some Northern Sotho speech tasks. Tonal languages, such as Northern Sotho, may have larger f_0 ranges than non-tonal languages. Producing the lexical tones that characterize tonal languages requires greater f_0 modifications than producing intonation, as in non-tonal languages, alone [27]. The mean f_0 was

significantly higher during the Northern Sotho reading task than during the Northern Sotho monologue. This suggests that f_0 variations do not manifest in the same manner across languages even when identical elicitation tasks are used. Clear f_0 contrasts within the Northern Sotho samples indicate that the tonal language may be more susceptible to f_0 variations in specific speech tasks than non-tonal languages; in this case, English. A small number of cross-linguistic bilingual studies have investigated tonal languages [5, 6]. Of these studies, fewer investigated vocal characteristics across differing speech tasks [4-7]. Continued cross-linguistic research is needed to corroborate these findings with similar studies.

All the participants in the current study were sequential bilinguals, and their acquisition of Northern Sotho preceded their acquisition of English. Many South Africans acquire English at school and continue speaking a tonal language at home [10]. Acquiring English as an academic language necessitates reading regularly in English [23]. It is likely that the participants were more familiar with reading aloud in English than in Northern Sotho as they would have had to regularly read aloud in English for academic purposes. Within the English samples, the English reading passage elicited a significantly quicker articulation rate than both the picture description and monologue tasks. Reading aloud may underscore a language's acoustic properties as words are inadvertently hyperarticulated throughout the structured task [6, 19, 33]. The nature of reading aloud, paired with the participants' familiarity with reading aloud in English, may have contributed to the English reading task eliciting significantly quicker articulation rates within the English tasks.

The English passage reading appeared to mediate significantly quicker articulation rates both within the English speech samples and across the English-Northern Sotho passage readings. The English-Northern Sotho female bilinguals read significantly quicker in English than they did in

Northern Sotho. This result contrasts with English-Mandarin and English-Korean bilingual females who read significantly slower in English [6]. Articulation rates are thought to be quicker in tonal languages, such as Northern Sotho, where less semantic information is packed into each syllable [6]. On the other hand, English is characterized by high-information density syllables with slow syllabic rate [6]. English was the participants' non-native language in both the current study and in Lee and Sidtis' [6] study. However, approximately 65% of South Africans are taught in English throughout primary school [40]. The quicker articulation rate evoked by the English reading passage in the current study may be ascribed to the participants' familiarity with reading in English given the language's role in South African education. This suggests that the higher mean f_0 in the English reading task, both within- and across-languages, manifested as a function of this specific sample's language acquisition history and daily language use.

Most of the participants reported conversing in Northern Sotho at home with family. The participants may have had less experience reading aloud in Northern Sotho. Performing a relatively unfamiliar speech task, while being audio-recorded, could have caused some degree of stress that contributed to elevated mean f_0 in the Northern Sotho reading task compared with the Northern Sotho monologue [27]. Mean f_0 differences within the Northern Sotho samples suggests that the task-effect might be more transparent in tonal languages than in non-tonal languages, such as English. The current study proposes that the bilingual's daily language use, and acquisition history, provides valuable context within which to interpret the interaction between language, voice, and speech tasks.

The monologue task was thought to best-represent natural communication [35]. No significant acoustic or perceptual differences manifested across the English and Northern Sotho monologues. Language-based acoustic fluctuations were most prevalent during the reading task.

When the reading task was compared with the monologues within each language, it was found that intensity and mean f_0 differed uniquely in English and Northern Sotho respectively. The bilinguals read with a significantly quicker articulation rate in English yet read with a significantly higher mean f_0 in Northern Sotho. Contrasting structured and unstructured speech tasks highlights the unique way speech tasks interacted with each of the bilinguals' languages. SLPs may take cognizance of possible language and task effects in practice by assessing and treating bilingual voice clients using multiple structured and unstructured elicitation tasks in more than one language.

In conclusion, the language spoken by an individual may be an acquired influence on voice. Differing manifestations of acoustic voice parameters across and within English and Northern Sotho samples suggest that different languages do not interact with speech tasks in a universal manner. Acoustic analysis may reveal vocal characteristics that are neither perceptually obvious to the SLP nor evident in only one of the bilingual's languages. Isolating language as an acquired influence on voice remains challenging as there are limited opportunities to control for proficiency in linguistically-diverse contexts. A potential drawback of the current study is that proficiency in English and Northern-Sotho was based on the participants' self-reports as a valid language proficiency screening tool was not available for the English-Northern Sotho bilingual populace. The bilingual speaker's language acquisition history and daily use may inform interpretation of differing vocal characteristics across languages and speech tasks. SLPs require further insight into the cross-linguistic effect on voice is to inform best-practice in clinical assessment, treatment and decision-making when working with bilingual voice clients.

Acknowledgements

The authors extend appreciation to Prof. Elsabé Taljard for kindly reviewing the Northern Sotho Rainbow Passage.

Statement of Ethics

This study protocol was reviewed and approved by the Research Ethics Committee of the Faculty of Humanities at the University of Pretoria (HUM018/0221). Written informed consent was obtained from each participant.

Declaration of Interest

The authors report there are no competing interests to declare.

Funding Sources

There are no funding sources to declare.

Author Contributions

AH conducted data collection and was the main contributor in writing the manuscript.

BP conceptualised the method, contributed to interpretation of results, and edited the manuscript.

MG conducted the statistical analysis and assisted with the presentation of results.

JvdL conceptualised the method, contributed to interpretation of results, and edited the manuscript.

Data Availability Statement

Data will be published and managed in an institutional Research Data Management system; an accredited open data repository. Further enquiries can be directed to the corresponding author.

References

1 Van der Merwe A, Le Roux M. Idiosyncratic sound systems of the South African Bantu languages: Research and clinical implications for speech-language pathologists and audiologists.

2 Ferrand CT. Voice disorders: Scope of theory and practice. Pearson Higher Ed; 2011

3 Godin KW, Hansen JH. Physical task stress and speaker variability in voice quality. EURASIP Journal on Audio, Speech, and Music Processing. 2015 Dec;2015(1):1-3.

4 Nevo L, Nevo C, Oliveira G. A comparison of vocal parameters in adult bilingual Hebrew-English speakers. InCoDAS 2015 Sep (Vol. 27, pp. 483-491). Sociedade Brasileira de Fonoaudiologia.

5 Zhu S, Chong S, Chen Y, Wang T, Ng ML. Effect of Language on Voice Quality: An Acoustic Study of Bilingual Speakers of Mandarin Chinese and English. Folia Phoniatrica et Logopaedica. 2022;74(6):421–30.

6 Lee B, Sidtis DV. The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. Speech, Language and Hearing. 2017 Jul 3;20(3):174-85.

7 Pépiot E, Arnold A. Cross-Gender Differences in English/French Bilingual Speakers: A Multiparametric Study. Perceptual and Motor Skills. 2021 Feb;128(1):153-77.

7 Pineda-Pérez E, Calvache C, Cantor-Cutiva LC. Bibliometric Analysis and Review of Literature on the Relationship Between Voice Production and Bilingualism. Journal of Voice. 2021 Aug 14.

8 Ansaldo AI, Saidi LG. Aphasia Therapy in the Age of Globalization: Cross-Linguistic Therapy Effects in Bilingual Aphasia. Behavioural Neurology. 2014; 2014:1–10.

10 Posel D, Hunter M, Rudwick S. Revisiting the prevalence of English: Language use outside the home in South Africa. Journal of Multilingual and Multicultural Development. 2020 Jun 9:1-3.

11 Graham C. Fundamental frequency range in Japanese and English: The case of simultaneous bilinguals. Phonetica. 2014;71(4):271-95.

12 Ordin M, Mennen I. Cross-linguistic differences in bilinguals' fundamental frequency ranges. Journal of Speech, Language, and Hearing Research. 2017 Jun 10;60(6):14103-506.

13 Ng ML, Chen Y, Chan EY. Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers—a long-term average spectral analysis. Journal of Voice. 2012 Jul 1;26(4): e171-6.

14 Cantor-Cutiva LC, Bottalico P, Nudelman C, Webster J, Hunter EJ. Do voice acoustic parameters differ between bilingual english-spanish speakers and monolingual english speakers during english productions? Journal of Voice. 2021 Mar 1;35(2):194-202.

15 Bu-Al-Makarem A. Reading and Spontaneous Speaking Fundamental Frequency Of Young Arabic Men For Arabic And English Languages: A Comparative Study. Perceptual And Motor Skills. 2007;105(6):572.

16 Järvinen K, Laukkanen AM, Geneid A. Voice Quality in Native and Foreign Languages Investigated by Inverse Filtering and Perceptual Analyses. Journal of Voice [Internet]. 2017 Mar 1 [cited 2021 Sep 26];31(2): 261.e25–31.

17 Ryabov R, Malakh M, Trachtenberg M, Wohl S, Oliveira G. Self-perceived and Acoustic Voice Characteristics of Russian-English Bilinguals. Journal of Voice. 2016 Nov;30(6): 772.e1–8.

18 Levi SV. Another bilingual advantage? Perception of talker-voice information. Bilingualism: Language and Cognition. 2018 May;21(3):523-36.

19 Hudson AI, Holbrook A. Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading. Journal of Speech, Language, and Hearing Research. 1982 Mar;25(1):25-8.

19 McAllister A, Brandt SK. A Comparison of Recordings of Sentences and Spontaneous Speech: Perceptual and Acoustic Measures in Preschool Children's Voices. Journal of Voice. 2012 Sep;26(5): 673.e1–5.

20 Khokhlova I. Lingua Franca English of South Africa. Procedia-Social and Behavioral Sciences. 2015 Dec 5; 214:983-91.

22 Lewis MP. Ethnologue: Languages of the world. SIL international; 2009.

23 Mariette K. Census 2011 Census in brief [Internet]. 2012. Available from: http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

24 Louw C, Swanepoel DW, Eikelboom RH, Myburgh HC. Smartphone-based hearing screening at primary health care clinics. Ear and Hearing. 2017 Mar 1;38(2): e93-100.

25 Roberts, P., & Shenker, R. Assessment and treatment of stuttering in bilingual speakers. In E.G. Conture, & R. F. Curlee (Eds.), Stuttering and related disorders of fluency (pp. 183-209). Thieme Medical Publishers, Inc; 2007

26 Kohnert K. Language disorders in bilingual children and adults. San Diego, Ca ; Oxford ; Melbourne: Plural Publishing; 2013.

27 Keating P, Kuo G. Comparison of speaking fundamental frequency in English and Mandarin. The Journal of the Acoustical Society of America. 2012 Aug;132(2):1050-60.

28 Altenberg EP, Ferrand CT. Perception of Individuals with Voice Disorders by Monolingual English, Bilingual Cantonese–English, and Bilingual Russian–English Women. Journal of Speech, Language, and Hearing Research. 2006 Aug;49(4):879–87.

29 Fairbanks G. Voice and articulation drill book (Second). Harper & Brothers; 1960

30 Goodglass H, Caplan E. The assessment of aphasia and related disorders. Lea & Febiger; 1972

31 Yamauchi EJ, Imaizumi S, Maruyama H, Haji T. Perceptual evaluation of pathological voice quality: A comparative analysis between the RASATI and GRBASI scales. Logopedics Phoniatrics Vocology. 2010 Oct 1;35(3):121-8.

32 Boersma P, Weenink D. Praat: Doing phonetics by computer (Version 5.4.560) [Computer program]. Available from: http://www.praat.org/

33 Cheng A. Cross-linguistic pitch differences in bilingual speakers. The Journal of the Acoustical Society of America. 2019 Mar;145(3):1825-.

34 Robb MP, Maclagan MA, Chen Y. Speaking rates of American and New Zealand varieties of English. Clinical Linguistics & Phonetics. 2004 Jan;18(1):1–15.

35 Kreiman J, Sidtis D. Foundations of voice studies: An interdisciplinary approach to voice production and perception. John Wiley & Sons; 2011 Mar 21.

36 de Winter JFC, Dodou D. Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). Practical Assessment, Research, and Evaluation [Internet]. 2019 Nov 23 [cited 2021 Apr 27];15(1). Available from: https://scholarworks.umass.edu/pare/vol15/iss1/11/

37 Gibbons JD, Subhabrata Chakraborti. Nonparametric Statistical Inference. 2010.

38 Field A. Discovering Statistics Using IBM SPSS Statistics. Pflege. 2014 Dec;27(6):430–0.

39 Tilsen S. A shared control parameter for F0 and intensity. In: Speech Prosody 2016. ISCA: ISCA; 2016.

40 Heugh K. Multilingual education policy in South Africa constrained by theoretical and historical disconnections. Annual Review of Applied Linguistics. 2013 Mar; 33:215-37.

Table 1 *The participants' estimated ages of acquisition of English and Northern Sotho and current daily environmental use of both languages*

Acquisition and daily use	English		Northern Sotho	
	n	%	n	%
Period of acquisition				
First language	0	0	17	89.5
Preschool	11	57.9	1	5.3
Primary school	8	42.1	1	5.4
Daily environmental use				

Home	6	31.6	18	94.7
Work	9	47.4	10	52.6
Social	14	73.7	13	68.4
Daily communication partners				
Family	6	31.6	17	89.5
Co-workers	15	78.9	8	42.1
Friends	17	89.5	13	68.4

Table 2 *Voice analysis protocol adapted from Lee and Sidtis (2017) and Nevo et al. (2015)*

Analysis	Apparatus	Speech Task(s)	Parameters

Acoustic	Praat software (v.5.4.56) [32]	Passage Reading Picture description Monologue	Mean f_0 Intensity Articulation rate
Perceptual	GRBASI scale [33] Nominal scale	Monologue	Voice quality Resonance Glottal attack

Table 3 *Mean, standard deviation (SD), median, and interquartile range of the mean f0, intensity, and speech rate across English and Northern Sotho speech tasks*

Acoustic parameter	English			Northern Sotho			WSR	p-value
	Mean (SD)	Median	IQR	Mean (SD)	Median	IQR		
Fundamental Frequency (Hz)	199.57	196.53	19.6	203.83	197.71	25.88	-1.29	0.21

Passage Reading	(-21.79)			(-19.60)				
Picture description	196.50	192.81	34.75	204.61	199.11	34.06	-2.94	0.002*
	(-23.18)			(-19.49)				
Monologue	194.87	190.47	35.36	191.11	184.56	23.17	-1.29	0.21
	(-24.83)			(-21.70)				
Overall	196.98	194.87	21.01	199.85	194.16	21.39	-1.65	0.10
	(-22.30)			(-18.70)				
Intensity (dB)								
Reading passage	65.09	64.23	4.06	66.38	65.34	5.59	-2.11	0.03*
	(-4.14)			(-4.28)				
Picture description	65.01	62.88	3.71	65.88	63.53	7.71	-1.65	0.10
	(-5.00)			(-4.70)				
Monologue	64.38	62.55	5.29	63.79	61.96	7.89	-0.97	0.35
	(-5.44)			(-5.49)				
Overall	64.83	63.13	4.09	65.35	63.31	5.63	-1.15	0.27

	(-4.57)			(-4.35)				
Articulation rate (syllables/sec)								
Reading passage	3.80 (-0.30)	3.8	0.36	3.40 (-0.51)	3.46	0.81	-2.92	0.002*
Picture description	3.40 (-0.52)	3.43	0.91	3.67 (-0.421)	3.71	0.41	-1.77	0.08
Monologue	3.46 (-0.30)	3.42	0.53	3.68 (-0.41)	3.66	0.53	-1.85	0.07
Overall	3.55 (-0.32)	3.60	0.54	3.58 (-0.32)	3.69	0.51	-0.04	0.98

*Statistically significant ($p < 0.05$)

Table 4 *Task-based acoustic differences within English and Northern Sotho*

Language	Acoustic parameter	Pairwise comparison	Pairwise Z-test	p-value

English	Articulation rate	description - reading	3.49	<0.001*
		reading - monologue	3.08	0.002*
		monologue - description	-0.41	0.68
Northern Sotho	f_0	description - reading	0.00	1.00
		reading - monologue	3.41	<0.001*
		monologue-description	3.41	<0.001*

*Statistically significant ($p < 0.05$)