THE LINGUAL VOICE QUALITY SETTINGS OF STANDARD SINGAPORE ENGLISH AND SINGAPORE COLLOQUIAL ENGLISH

Lin Jingmin & Scott R. Moisik

Nanyang Technological University
jingmin001@e.ntu.edu.sg; scott.moisik@ntu.edu.sg

ABSTRACT

This paper investigates the voice quality settings of the tongue in the two registers of Singapore English, Standard Singapore English (SSE) and Singapore Colloquial English (SCE). Visualisation of the tongue was achieved through lingual ultrasound, and tongue imaging data was gathered from 7 (5 Singaporean, 2 non-Singaporean) female undergraduates during SSE and SCE speech. Tongue contours during interspeech postures (ISPs) from each register were statistically evaluated with Generalised Additive Mixed Modelling (GAMM). Additionally, raw ultrasound data was subjected to Principal Components Analysis (PCA), and randomisation was used to test for the effect of register on mean PC scores. Results indicate that SCE is characterised by a significantly lower tongue tip than in SSE. These findings may have important implications on the phonetic segments of each register and serve as a promising first step into voice quality research of Singapore English.

Keywords: Singapore English, lingual ultrasound, voice quality, voice quality settings

1. INTRODUCTION

The main objective of this paper is to examine the voice quality settings of the two registers of Singapore English, namely, Standard Singapore English (SSE) and Singapore Colloquial English (SCE). More specifically, this paper aims to identify, describe, and assess any systematic, register-specific differences in the voice quality settings of the tongue by means of lingual ultrasound. Knowledge of these voice quality settings will improve our understanding of the Singapore English accent and inform us about potential interactions with its segmental structure.

2. BACKGROUND

2.1. Voice quality and voice quality settings

The term ‘voice quality’ has two main senses in the field of phonetics; the narrow sense is concerned only with phonation type (implicating laryngeal vibration). For the purposes of this paper, ‘voice quality’ refers to its broader sense, defined by Abercrombie [1] as “the quasi-permanent quality of a speaker’s voice”, which also includes supralaryngeal features.

Laver [14] postulates that voice quality mainly derives from two sources: (1) the anatomical and physiological foundation of the speaker’s vocal apparatus; and (2) the long-term tendency of the speaker to use certain muscular settings of his vocal apparatus. The latter, also known as voice quality settings, presumably gives rise to the ‘auditory colouring’ of an individual’s voice, and, on a larger scale, characterizes the accent, or overall ‘sound’, of the individual’s language variety [6].

In fact, several studies have empirically observed the existence of language-specific voice quality settings. For instance, differences have been observed between the settings of English and other major languages, such as French [8, 27], Polish [22], and German [2]. In addition, differences in voice quality settings have also been detected between two Dutch dialects [27]. However, while research into voice quality has garnered a modest amount of attention in recent years, major language varieties, such as Singapore English, remain wholly neglected in this subfield of phonetics.

2.2. Voice quality settings of SSE and SCE

Singaporeans switch between SSE and SCE mainly depending on register, but speakers also use these subvarieties to index particular social characteristics associated with each [4]. SSE and SCE vary mainly in terms of morphology and syntax [10], but they also *sound* different from one another, each possessing a rather distinct phonetic and phonological profile. For example, SSE is more ‘diphthongal’ [15], has longer VOTs in its voiceless stops [17], and has a high incidence of postvocalic-r [24]. SCE, on the other hand, has a high tendency of word-final dark [ɫ] vocalisation [23].

Based on the different phonetic realisations that each register has, it is therefore hypothesised that the underlying lingual voice quality settings of SSE and SCE will be significantly different from each other.

3. METHODOLOGY

3.1. Participants

Seven participants were recruited by word of mouth, comprising of 5 Singaporeans, 1 Malaysian, and 1 Chinese. The latter two have had 4 and 12 years of education in Singapore, respectively. All participants were female, between the ages of 18 and 24, ethnic Chinese, and undergraduates at NTU. It was ideal to control for any gender, age and ethnic effects, and to ensure that all participants had a good command of both SCE and SSE—the latter of which is often claimed to be spoken only by the educated Singaporean population [9].

3.2. Procedure: Spontaneous and Read Speech

In order to maintain some level of ecological validity, especially in the case of SCE, it was most ideal to examine conversational speech. Thus, participants went through two spontaneous dialogues: first with a university professor whom participants did not know personally, on topics relating to school (for SSE), and next with a friend, on topics relating to leisurely activities (for SCE). This method was previously successfully employed by Moorthy & Deterding [19] to elicit a change in register in their participants.

The main drawback with spontaneous speech is that segmental content is uncontrolled, and variation in the phonemic contexts surrounding the pause may confound with any signal arising purely from voice quality [18]. Therefore, in addition to the spontaneous task, participants were instructed to read the short version of the 'Rainbow passage' twice in a 'proper' and 'standard' manner (for SSE) and twice in 'Singlish' (for SCE). The 'Rainbow passage' [7] was chosen as it is a phonetically balanced passage such that the ratios of the various phonemes reflect that of normal, unscripted speech.

3.3. Apparatus

A MC10-5R10S-3 element (10 mm, 5–10 MHz) microconvex ultrasound probe (operated using the SonoSpeech micro ultrasound system by Articulate Instruments) was placed under the submental surface of the chin to obtain a sagittal image of the tongue, stabilised by means of a lightweight aluminium helmet worn by participants throughout the experiment. Both the probe and an AT3035 Audio-Technica cardioid condenser microphone (plugged into a Focusrite Scarlett Solo audio interface using 16-bit sampling at 44.1 kHz) were connected to a computer, where the accompanying AAA software automatically captured and synchronised ultrasound videos and audio signals produced by participants.

3.4. Analysis 1: Interspeech posture (ISP)

Gick et al. [8] suggests that a language's voice quality setting can be inferred from its *interspeech posture* (ISP), a "motionless state of the articulators" which in turn can be derived from an *interutterance pause* [28]. As the definition of a pause has been quite inconsistent in the existing literature [8, 20, 26, 27], the present study took an all-encompassing approach, to include all pauses occurring between speech, including inhalation, grammatical pauses (e.g. commas, periods), and ungrammatical pauses (e.g. hesitation, word-search). The only pauses that were definitively excluded were those associated with swallowing.

A grand total of 710 pauses across all participants were manually segmented and annotated based on the acoustic signal using Praat [3]. Subsequently, the auto-synchronised ultrasound frames corresponding to each pause were extracted using MATLAB R2018b, where the center frame of a pause was selected as the target frame from which the ISP can be obtained. The center frame was preferred over the mean frame (across the duration of the pause) to minimise carryover co-articulatory effects from neighbouring words as much as possible [27].

The ISPs, i.e., the tongue contours, in the target frames were manually traced using MATLAB R2018b. Following which, each trace was resampled into 200 x- and y- coordinates, from which a mean trace was constructed. This mean trace served as the reference trace for a Generalised Procrustes Analysis (GPA), which improves (but does not fully guarantee) biological homology across all traces. Once all the traces were superimposed and registered with GPA, the resulting x- and y- coordinates were used as the basis of the data analysis.

The data were fit into generalized additive (mixed) models (GAM) using the `bam()` function from the R package `mgcv` [29]. The independent variable of register (SCE vs. SSE) was incorporated as a difference smooth, and participants were treated as random smooths.

The main challenge with GAM relates to significance testing, which is less straightforward as compared to linear models. Although there are multiple methods to significance testing, not all are adequate or conservative enough. Sóskuthy [21] suggests that the "most reliable and least anti-conservative" procedure for GAM significance testing is to first conduct a model comparison using the `compareML()` function from the R package `itsadug` [25]. This allows us to observe if there is an overall significant difference in tongue contours between registers. Subsequently, we should plot (A) the predicted tongue contours with corresponding pointwise confidence intervals, as well as (B) the difference smooth (excluding random effects) along with a confidence interval. These will provide us with graphical interpretations of the results.

3.5. Analysis 2: PCA on raw ultrasound data

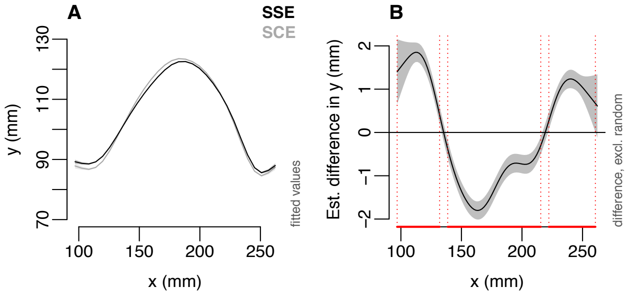
In complement to ISP analysis, the raw ultrasound data for the read speech conditions from both registers were subjected to Principal Components Analysis (PCA) within participants (minimally four videos, two repetitions for each register), treating each video frame as an observation and each pixel as a variable. Thus, the analysis applies to all events in a given video, not just the ISP. This allows for information not just about the tongue contour but also the entire appearance of the oral structures in the ultrasound video to be taken into consideration.

4. RESULTS

4.1. Analysis 1: ISP differences between registers

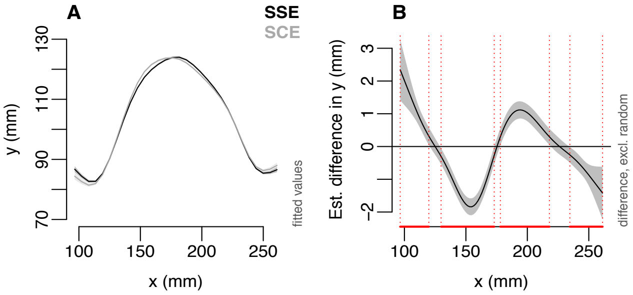
Under the spontaneous speech condition, the data were fit using two GAMs (a nested one without register, and a full one with register). Model comparison reveals that the tongue contours between registers differ significantly overall (nested [score = 283644.0, edf = 5] vs. full [score = 283188.5, *edf* = 8] diff. = 455.49, *df* = 3, *p*-value < 2e–16). We can see from Fig. 1 that the areas with significant differences along the tongue contours correspond roughly to the tongue tip/blade, tongue body, and tongue root. Specifically, SSE is characterised by a higher tongue tip, lower tongue body, and a slightly more retracted tongue root.

Figure 1: Predicted tongue contours (A) and difference smooth with confidence interval (B) under the spontaneous speech condition. The tongue tip is at the lower values of x. A positive value on the y-axis (in B) indicates that the contour is higher in SSE.



For the read speech condition, the data were also fit using two GAMs (a nested one without register, and a full one with register). Model comparison reveals that the tongue contours between registers differ significantly overall (nested [score = 150908.5, *edf* = 5] vs. full [score = 150728.1, edf = 8] diff. = 180.47, *df* = 3, *p*-value < 2e–16). Fig. 2 shows that the areas of significant differences correspond roughly to the tongue tip/blade, tongue body, and tongue root. Once again, SSE possesses a higher tongue tip. However, it appears difficult to draw any systematic conclusions about the tongue body and tongue root.

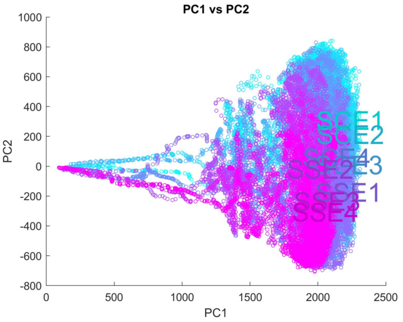
Figure 2: Predicted tongue contours (A) and difference smooth with confidence interval (B) under the read speech condition. The tongue tip is at the lower values of x. A positive value on the y-axis (in B) indicates that the contour is higher in SSE.



4.2. Analysis 2: PCA differences between registers

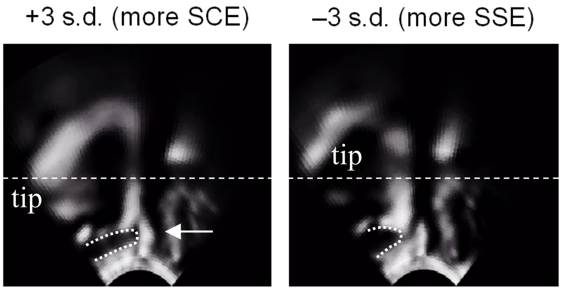
Fig. 3 illustrates the PCA of the read speech condition for one participant, focusing on PC1-2 space: each point represents one video frame and all frames were used in the analysis. Frames close together in this space are similar in their brightness pattern. Mean-warping inspection of the PCs revealed that PC1 consistently detected a brightness ramp-up (which is evident when recording is started with the AAA software), thus it was not included in statistical analysis.

Figure 3: PCA for participant P100301 (a native Singaporean) showing frames from eight ultrasound videos of the Rainbow passage (four for SSE and four for SCE) as points in PC1-2 space.



Randomization testing with restricted permutations (within participants) on the overall mean of mean PC2-20 scores (taken across all frames within a given trial) was used to gauge whether there were any differences between registers. With 20K permutations, the observed difference in means has a *p*-value of 0.025, thus allowing for rejection of the null hypothesis that the ultrasound appearance of the tongue (in PC2-20) does not differ between registers.

Figure 4: Mean warps (at +/–3 s.d.) along PC2 (2.7% of the variation) of raw ultrasound data of participant P100301. Left side is anterior.


+3 s.d. (more SCE)

–3 s.d. (more SSE)

Fig. 4 shows PC2 warping applied to the mean frame at +/–3 s.d., with higher values being more like SCE and lower values being more associated with SSE (cf. Fig. 3). Visually, the tongue tip appears to be slightly lower in the warp associated with SCE (left side, compare using the dashed white line), which is consistent with the results from the ISP analysis. Another difference (indicated by the white arrow) is the slackness of the geniohyoid muscles (dotted white line) and resulting posterior hyoid positioning (white arrow) in SCE.

5. DISCUSSION

5.1. Potential interaction between voice quality settings and phonetic segments in SCE and SSE

The results demonstrate that the tongue tip is consistently lower in SCE than SSE across both conditions. Since no systematic pattern was observed for the tongue body and tongue root, the study shall refrain from making any premature conclusions about them. The discussion will focus on the tongue tip.

Prior to that, it should be highlighted that the ‘chicken-and-egg’ problem persists: it is unclear whether the long-term, habitual use of certain muscular settings results in the realisation of phonetic segments, or vice-versa (or even perhaps both). For the sake of argument, the paper follows Laver’s [12] proposition, in that voice quality settings may best be exemplified as a form of articulatory ‘bias’, where a setting is a “constraining influence on segmental action”.

Thus, that SCE has a lower tongue tip may have a profound effect on certain segments where the tongue tip is the active articulator [11], including /r, l, θ, ð/. If Singaporeans habitually maintain a lower tongue tip in SCE, it may relate, to some extent, as to why word-final dark [ɫ] in SCE is often vocalised, substituted (with [w]), and sometimes even deleted altogether [5, 16, 23]. Conversely, a higher tongue tip in SSE may be associated with more frequent usage of postvocalic-r in SSE [24], as coda /r/ is known to have a high degree of coarticulatory aggression, and may exert long distance coarticulation effects [26].

5.2. Limitations

A main limitation is the lack of randomisation of tasks—all participants were put through the SSE tasks before the SCE tasks. This sequence was deemed necessary as participants were uncomfortable using SCE right at the beginning of the experiment. The problem is that there could potentially be “probe-shifting-over-time” effects, i.e. differences observed might be due to slight changes in the position of the probe as the experiment progressed. It should be noted however that the voice quality setting itself might influence the positioning of the probe because of differences in muscular tensioning, particularly of the mylohyoid and geniohyoid muscles (as is evident in the PCA, see Fig. 4).

Another issue, which is a problem incidental to ultrasound imaging in general, relates to normalizing tongue contours across speakers [18]. Although the study endeavoured to normalise all tongue contours using GPA, it by no means guarantees biological homology between speakers.

6. CONCLUSION

Using lingual ultrasound, the study examined the voice quality settings of the tongue between SCE and SSE. Evidence from the interspeech posture analysis and PCA of the raw ultrasound video signal support the conclusion of (minimally) a difference in tongue tip height between the registers, but caution still is required in the interpretation as the effects may be associated with probe-shifting or segmental influences (due to co-articulation in the ISP or directly in the PCA). So, more work is required to validate the results. Future studies could also extend the research by looking at other aspects of voice quality settings in Singapore English, such as the lips, jaws and larynx.

7. REFERENCES

[1] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh University Press.

[2] Benítez, A., Ramanarayanan, V., Goldstein, L., Narayanan, S. S. 2014. A real-time MRI study of articulatory setting in second language speech.

In *Fifteenth Annual Conference of the International Speech Communication Association*.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43. *http://www.praat.org/*

[4] Cavallaro, F., Ng, B. C. 2009. Between status and solidarity in Singapore. *World Englishes* 28, 143-159.

[5] Deterding, D. 2007. Phonetics and Phonology. In: Deterding, D. *Singapore English.* Edinburgh: Edinburgh University Press, 12-39.

[6] Esling, J. H., Wong, R. F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL quarterly* 17, 89-95.

[7] Fairbanks, G. 1960. The rainbow passage. *Voice and articulation drillbook* 2.

[8] Gick, B., Wilson, I., Koch, K., Cook, C. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61, 220-233.

[9] Gupta, A. F. 1989. Singapore Colloquial English and Standard English. *Singapore Journal of Education* 10, 33-39.

[10] Gupta, A. F. 1994. *The step-tongue: Children's English in Singapore* 101. Multilingual Matters.

[11] Honikman, B. 1964. Articulatory Settings. In: Abercrombie, D., Fry D. B., MacCarthy P.A.D.,. Scott N.C, Trim, J.L.M. (Eds.), *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961.* London: Longmans, Green & Co. Ltd, 73-84.

[12] Laver, J. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.

[13] Laver, J. 1994. *Principles of phonetics*. Cambridge University Press.

[14] Laver, J. D. M. 1968. Voice quality and indexical information. *British Journal of Disorders of Communication* 3, 43-54.

[15] Lee, E. M., & Lim, L. 2000. Diphthongs in Singaporean English: Their realisations across different formality levels, and some attitudes of listeners towards them. In: Brown, A., Deterding D., Low E. L. (Eds.), *The English language in Singapore: Research on pronunciation.* Singapore: Singapore Association for Applied Linguistics, 101-111.

[16] Lian, H. W. 2008. Phonological patterns in the Englishes of Singapore and Hong Kong. *World Englishes* 27, 480-501.

[17] Liu, P. Z. 2011. Voice onset time production in Singapore English. *The Journal of the Acoustical Society of America* 130, 2520-2520.

[18] Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S., Schaeffler, F. 2010. Measuring language-specific phonetic settings. *Second Language Research* 26, 13-41.

[19] Moorthy, S. M. & Deterding, D. 2000. Three or tree? Dental fricatives in the speech of educated Singaporeans. In: Brown, A., Deterding D., Low E. L. (Eds.), *The English language in Singapore: Research on pronunciation.* Singapore: Singapore Association for Applied Linguistics, 76-83.

[20] Ramanarayanan, V., Byrd, D., Goldstein, L., Narayanan, S. S. 2010. Investigating articulatory setting-pauses, ready position, and rest-using real-time MRI. In *Eleventh Annual Conference of the International Speech Communication Association.*

[21] Sóskuthy, M. 2017. Generalised Additive Mixed Models for dynamic analysis in linguistics: A practical introduction. arXiv:1703.05339

[22] Święciński, R. 2013. An EMA study of articulatory settings in Polish speakers of English. In: *Teaching and researching English accents in native and non-native speakers.* Berlin, Heidelberg: Springer, 73-82.

[23] Tan, K. K. 2005. Vocalisation of /l/ in Singapore English. In Deterding, D., Brown, A. & Low E. L. (Eds.), *English in Singapore: Phonetic Research on a Corpus*. Singapore: McGraw-Hill, 43-53.

[24] Tan, Y. Y. 2012. To r or not to r: Social correlates of /ɹ/ in Singapore English. *International Journal of the Sociology of Language* 218, 1-24.

[25] van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H. 2016. itsadug: Interpreting Time Series and Autocorrelated Data using GAMMs. R package version 2.2.

[26] West, P. 2000. Long-distance coarticulatory effects of British English /l/ and /r/: An EMA, EPG and acoustic study. In *Proccedings of the 5th Seminar on Speech Production: Model and Data*, 105-108.

[27] Wieling, M., Tiede, M. 2017. Quantitative identification of dialect-specific articulatory settings. *The Journal of the Acoustical Society of America* 142, 389-394.

[28] Wilson, I., Gick, B. 2014. Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research* 57, 361-373.

[29] Wood, S. N. 2017. mgcv: mixed GAM computation vehicle with automatic smoothness. R package version 1.8-22.