Ultrasound Observation of Tongue Contour for Sibilant Fricatives in Fast Speech: A Pilot Study

Misato Shimizu[1], Takayuki Arai[1], Ai Mizoguchi[2, 3], Maho Morimoto[4, 1], Weiyu Li[1], Noriko Yamane[5]

[1]Sophia University, Japan, [2]Maebashi Institute of Technology, Japan, [3]NINJAL, Japan,
[4]Chuo University, Japan, [5]Hiroshima University, Japan
m-shimizu-9r6@eagle.sophia.ac.jp, arai@sophia.ac.jp, aimizoguchi@maebashi-it.ac.jp,
maho.morimoto.jp@gmail.com, yamanen@hiroshima-u.ac.jp

Keywords: sibilant fricative, alveolar fricative, alveopalatal fricative, fast speech, speech error

1. Introduction

Some adult native Japanese speakers self-report difficulties in normal speech despite having no medical history related to speech disorders [1], and these individuals are expected to make speech errors more frequently than other speakers who do not perceive speech difficulty. This study aims to understand how speech errors occur by investigating the articulation process of fast speech. Previous studies show that speech errors occur when phrases are repeated rapidly [2, 3, 4]. These studies focused on the repeated articulation of plosive sounds in /tɑp kɑp/. Dawson [4] used ultrasound imaging to observe the tongue and confirmed errors such as *intrusions* and *reductions*. Intrusions are defined as the appearance of a movement consistent with a competing gesture within a segment, whereas reductions are defined as the undershooting of the constriction target relative to non-errorful utterances [4]. This study investigates Japanese sounds /sa/ and /ɕa/, which children find difficult to pronounce [5]. Similar phenomena may occur in adults during the pronunciation of fricative sounds with similar places of articulation. We investigated the types of pronunciation errors and how tongue contours vary when repeatedly pronouncing the phrase /sa ɕa/ to a fast-paced metronome beat.

2. Methods

Three native Japanese speakers (M: two, F: one) participated in the experiment. Prior to the experiment, a speech questionnaire was administered to the participants. They were aware of their difficulty in articulation even though they did not report any history of speech and hearing disorders.

Speech experiments were conducted individually in a soundproof room. AAA [6] was used for recording. An ultrasound probe was fixed under the participant's chin using an Ultrafit headset [7]. To control the speech rate, clip-on bone conduction earphones were employed to present a metronome sound. The Beats Per Minute (BPM) was set to 480 and presented for 6 seconds. Metronomic sound was created using Audacity [8]. Participants were instructed to pronounce syllables presented in katakana on a monitor in synchronization with the metronome sound. Prompts comprised bi-syllabic phrases such as /sa ɕa/.

Audio recordings were extracted as wave files using AAA, and both audio and spectrograms were analyzed using Praat [9]. Five correct tokens and all error tokens (the number of error tokens varied among the speaker) were chosen from the entire speech based on the experimenter's auditory judgement of each speaker. Following Li et al. [5], the onset of the fricative was defined as the initial presence of aperiodic noise, and the offset was defined as the first zero-crossing of the periodic waveform of the following vowel. Annotations were created using Praat. Within these tokens, midpoints of /s/ and /ɕ/ were determined and labeled on TextGrids [9]. The TextGrid was imported into the AAA, and tongue contour tracking was conducted at these points. The five tongue contour frames of the correctly pronounced midpoints were averaged for each /s/ and /ɕ/ to be exported. For incorrect pronunciations, one frame of the midpoint was exported for each error. The contours of correct /s/ and correct /ɕ/ were overlaid, and all error contours were superimposed onto this image. Figures were created using the AAA Publisher [6].

3. Results

Figure 1 illustrates the study participants' tongue contours. The total number of errors recorded were as follows: SP01 had five errors, SP02 had four errors, and SP03 had four errors. Overall, there was a tendency for the tongue contour during the production of intended /ɕ/ to resemble /s/ rather than the reverse. This was observed for four tokens of SP01, two tokens of SP02, and two tokens of SP03, where the tongue contours resembled those of /s/ (Error Pattern 1). In the next most common case, the tongue was positioned somewhere between correct /s/ and correct /ɕ/ (Error Pattern 2). Each speaker had one token with this error. An error where the tongue contours for /s/ resembled /ɕ/ (Error Pattern 3) was observed in one token for SP02. Furthermore, a metathesis error where /s/ became /ɕ/ and /ɕ/ became /s/ (Error Pattern 4) was also confirmed once for SP03.

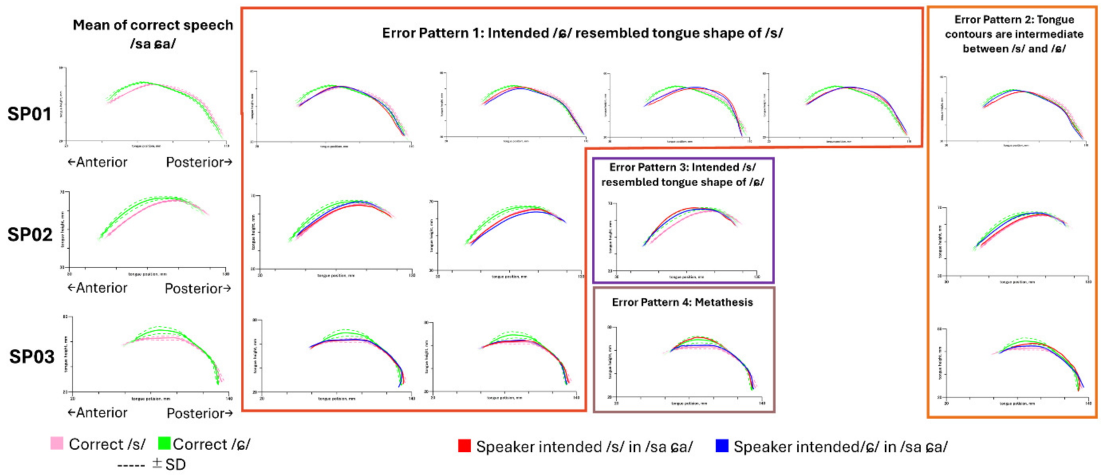

Figure 1: Overlay of the average tongue contour for accurate pronunciation and each token judged to be incorrect for the three speakers. The right side represents the anterior tongue, while the left side represents the posterior tongue. The axes were individually calibrated to accommodate each speaker's unique articulatory profile.

4. Discussion

In the three speakers, who were aware of their difficulties in speaking repeatedly pronounced /sa ɕa/ during fast speech, there was a tendency for the tongue contour of /ɕ/ to resemble /s/. This is shown in Pattern 1, where the tongue body is lowered so that the tongue contour of /ɕ/ overlaps with /s/, suggesting the possibility of reduction of the tongue dorsum. This is similar to the phenomenon reported in Dawson [4]. One of the potential reasons for the reduction is that the dorsum raising requires a larger muscular efforts compared to tongue tip raising. Conversely, there was only one instance of /s/ resembling /ɕ/ (Pattern 3), therefore intrusions of the tongue dorsum movement are less likely to occur.

Perhaps, increased attention to the initial sound contributed to the stability of /s/ pronunciation. However, as this experiment was conducted without changing the order of phonemes, this aspect requires further investigation. In future studies, more number of participants will be recruited and similar experiments will be conducted with different phoneme orders and combinations to elucidate the mechanisms of speech errors.

References

[1] Kitamura, T., Nota, Y., Hashi, M., Takemoto, H 2019. Survey of Japanese undergraduate and graduate students' awareness of clumsiness while speaking. *The Journal of the Acoustical Society of Japan*, 75(3), pp.118-124.

[2] Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., Byrd, D. 2007. Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412.

[3] Slis, A. 2014. The effect of articulatory constraints and auditory information on patterns of intrusions and reductions. *Doctoral Dissertation*. University of Toronto.

[4] Dawson, K. M. 2020. Speech Motor Control in Younger and Older Adults: The Effect of Age and Individual Differences on Speech Error Production. *Doctoral Dissertation*. The Graduate Center of the City University of New York.

[5] Li, F., Edwards, J., Beckman, M. E. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1), 111–124.

[6] Articulate Instruments Ltd. 2023. *Articulate Assistant Advanced (AAA)* Version 221.2.0.

[7] Spreafico, L., Pucher, M., Matosova, A. 2018. UltraFit: A Speaker-Friendly Headset for Ultrasound Recordings in Speech Science. *Interspeech*, 1517–1520.

[8] Audacity Team. 2023. *Audacity* ® Version 3.4.0.

[9] Boersma, P. and Weenink, D. 2023. *Praat: Doing phonetics by computer.* Version 6.3.15.