Latency analysis of speech shadowing reveals processing differences in Japanese adults who do and do not stutter

Rong Na A, Koichi Mori, Naomi Sakai

Research Institute of National Rehabilitation Center for Persons with Disabilities, Saitama, Japan

hohchahar-aruna@rehab.go.jp, mori-koichi@rehab.go.jp, sakai-naomi@rehab.go.jp

10.21437/Interspeech.2015-615

Abstract

Speech shadowing is a dual-task paradigm that could reveal certain features of speech processing, where the shadowing latency between the onsets of heard and reproduced speech has often been a key investigation tool. The present study investigated the shadowing latencies in native Japanese adults who do (AWS) and do not stutter (AWNS), with relevant analysis of speech errors. Fifteen AWS and fourteen AWNS participated in the study. They were required to shadow two meaningful Japanese passages of approximately 1.4 min. Fifty phrase onsets were chosen for measuring shadowing latencies. The resultant latencies were longer than previously reported for English in both groups, which most likely reflects the larger mean syllable numbers per word of Japanese than of English. The AWS group had a significantly shorter latency than the AWNS. Besides, it was significantly more error-prone than the AWNS. The eleven AWS whose latency was more than 500 ms showed a significant negative correlation (trade off) between speech errors and latencies, which was not the case with the AWNS group. Those results imply that only the AWS may have hit the limit of their working memory capacity during shadowing. Thus the shadowing paradigm brought new insights into speech processing and stuttering.

Index Terms: speech shadowing, latency, speech errors, stuttering

1. Introduction

Speech shadowing, or shadowing, is a speech repetition task in which one has to simultaneously listen to and repeat the same running speech with a delay as short as possible without seeing a transcribed text [1-5]. Speech shadowing has been applied to second language (L2) learning, because it improves the production of L2-specific prosodic features effectively [1-3]. It has also been used as a research tool for investigating certain features of speech processing [4-8]. Since its cognitive load is high due to the requirement for on-line parallel processing in audition, recognition, retention, vocalization and articulation of running speech, it could reveal various aspects of the cognitive processing during speech.

The shadowing latency or reaction time from the onset of heard to that of reproduced speech has been used as a key investigation tool. The mean latency of shadowing vowel-consonant-vowel (VCV, eg. /apa/, /ata/, /aka/) syllable stimuli was 213 ms to 291 ms [6-8]. Marslen-Wilson (1985) [5] used two 300 words meaningful passages to measure the latency of shadowing in normal speakers and found that the latency of “close-shadowers” (shorter latency participants) were less than 361 ms and that of “distant-shadowers” (longer latency participants) were more than 401 ms. In contrast, the shadowing latencies in L2 learners were longer than those of native English speakers irrespective of the language: The shadowing latencies of English-as-a-L2 learners (native in Japanese) were 638–1176 ms [9, 10], and those of Japanese-as-a-L2 learners (native in Chinese, Korean, Tagalog, or Indonesian) were 638-947 ms [11]. The shadowing latency depends on shadowing materials (text length and difficulty), individual proficiency and other yet-to-be-specified factors.

Shadowing also has been used in stuttering research, because it could induce fluency in adults who stutter (AWS) [12-15]. Stuttering is a speech disorder in which sounds, syllables, or part words are repeated or prolonged, and/or speech is involuntarily paused or blocked often [16-18]. AWS have shown abnormalities both in speech perception and production [16-18]. The previous studies using vowel stimuli resulted in the reaction time of AWS were significantly longer than that of adults who do not stutter (AWNS) (e.g. 270 *vs.* 236 ms, 340 *vs.* 279 ms) [19, 20, 21]. However, the shadowing latency of AWS has not been investigated using longer, meaningful sentences that required significant overlap of listening and repetition as used in the L2 training and in Marslen-Wilson (1985) [5]. One caveat interpreting the latency measurements is that the analysis of speech errors should also be incorporated because there may be a trade-off between the shadowing latency and the error rate.

Using the shadowing latency as a research tool, this study investigated (1) whether the speech processing during shadowing is in chunks of syllables or words, which should be clarified by comparing the latencies of native Japanese and English speakers, since Japanese words are composed of several syllables on average whereas English words have less than 2 syllables on average, (2) given the speech production impediments in AWS, whether they take a different strategy than AWNS in shadowing, and (3) if there is a difference in the strategy, what is a possible cause.

2. Method

2.1. Participants

Participants were fifteen AWS (twelve males, three females, mean age = 28.3 years, age range = 18-38 years) and fourteen AWNS (nine males, five females, mean age = 24.6 years, age range = 19-37 years), all of whom were native Japanese speakers. None of the participants had any speech, language, hearing, or neurological disorders except stuttering. The educational levels of the participants in the two groups were similarly distributed. Stuttering severity was assessed using the Japanese stuttering severity instrument [22] and the Overall Assessment of the Speaker’s Experience of Stuttering for Adults (OASES-A) [23] (Table 1). The stuttering severity assessment [22] included the core stuttering behaviors frequencies in oral reading, explanation of pictures, and a free talk (monologue) in the present study. The stuttering severity ratingswere defined as: normal, less than 3 %; very mild, 3-5 %; mild 5-12 %; moderate, 12-37 %; severe, 37-71 %; very severe, more than 71% of the averaged stuttering frequency in phrases (bunsetsu). OASES-A is a questionnaire to measures the impact of stuttering on a person's life, it collects information about the totality of the stuttering disorder, including: (a) general perspectives about stuttering, (b) affective, behavioral, and cognitive reactions to stuttering, (c) functional communication difficulties, and (d) impact of stuttering on the speaker's quality of life [24]. The degrees of stuttering impact in OASES were defined in five levels: mild, mild-to-moderate, moderate, moderate-to-severe, and severe.

Informed written consent was obtained from all the subjects prior to the experiment in accordance to the protocol approved by the National Rehabilitation Center for Persons with Disabilities Review Board.

Table 1. *Attributes of the AWS participants.*

Participant	Gender	Age	Stuttering severity rating	OASES impact rating
1	Male	18	Moderate	Severe
2	Female	37	Normal	Moderate
3	Male	38	Moderate	Mild-to Moderate
4	Male	22	Normal	Moderate
5	Female	19	Moderate	-
6	Male	29	Very mild	Moderate
7	Male	26	Normal	Severe
8	Male	37	Mild	Moderate
9	Female	37	Normal	Severe
10	Male	28	Very mild	Moderate
11	Male	34	Moderate	Moderate
12	Male	23	Mild	Moderate
13	Male	22	Moderate	Moderate
14	Male	23	Moderate	Moderate
15	Male	31	Normal	Moderate

-: Not tested

2.2. Materials

Two types of materials have been used to measure the shadowing latency in previous studies. One of them was nonsense words like vowel-consonant-vowel stimuli [6-8]. The other was meaningful phrases or passages [5, 25-26]. The present study chose the latter, since the former tends to generate results indistinguishable with simple repetition without overlapping hearing and speaking. Two audio materials consisting of 499 and 503 morae (114, 118 bunsetsu or phrases) [15] were used as model stimuli to be shadowed. Their original texts were beginner's Japanese as a L2, which were assumed to be easy enough for the native participants to shadow without seeing transcribed texts. The model speech had been recorded by one male native speaker of Japanese at a speaking rate of 6 morae per second (including pauses).

2.3. Procedure

Participants individually were seated in a sound-attenuated room and were required to shadow the model speech. During the shadowing task, they heard the model speech over headphones (HSC271, AKG) played back from a computer (PRECISION T5400, Dell). The model speech and participants' speech were recorded simultaneously on separate tracks with a multi-track recorder (DR-680, TASCAM) at the sampling rate of 48 kHz with a 24 bit A/D resolution.

2.4. Analyses

For the latency analysis, fifty measurement points (phrases or clauses) after a short pause were chosen at quasi-equal intervals from the two materials. The latencies between the same phrases in the model speech and the reproduction speech were measured only for correct and fluent responses. Figure 1 gives an example of the method to measure shadowing latencies using speech analysis software Praat [27].

The speech errors were measured consisting of word substitutions, word insertions, and word or part-word omissions. Dysfluencies specific to stuttering were not counted as speech errors, but omissions were counted even when they followed speech blocking. For this reason, an omission of continuous words or phrases were counted as one error. The percentage of speech errors per 100 phrases (bunsetsu) was calculated. Bunsetsu is a phrase unit of Japanese that comprises a content word followed by a function word.

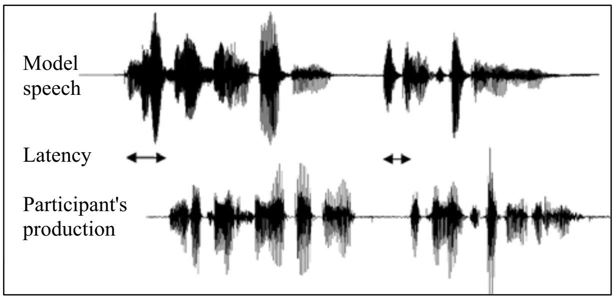

Figure 1: *Measurement of shadowing latencies. The upper waveform shows model speech, with the lower one participant's shadowed speech. The time lags (arrows) between the onsets of the same phrases in model and participant's speech are the shadowing latencies.*

3. Results

3.1. Shadowing latency in Japanese AWS and AWNS

The latencies of the AWS and AWNS groups were compared in Figure 2. The median latency of the AWS group was 727 ms, while that of the AWNS group was 914 ms. The shadowing

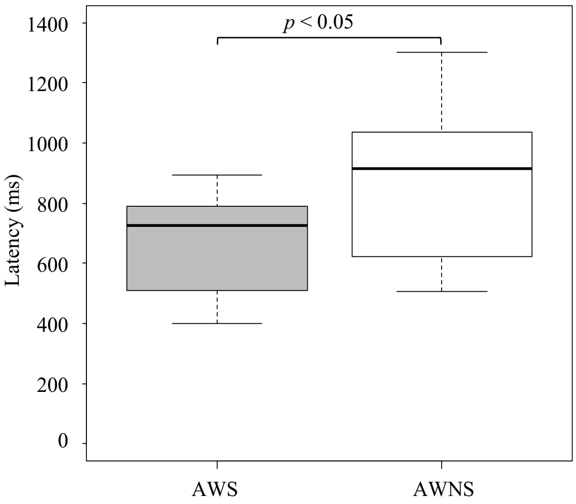

Figure 2: *Group comparison of shadowing latencies*

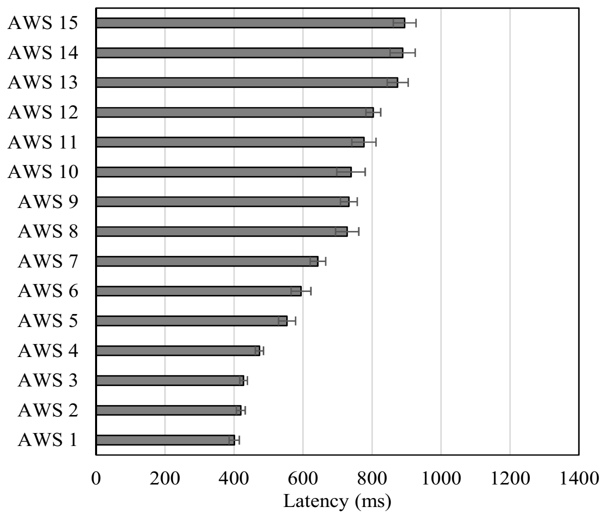

Figure 3: *Mean latencies of individual AWS. Error bars indicate standard errors.*

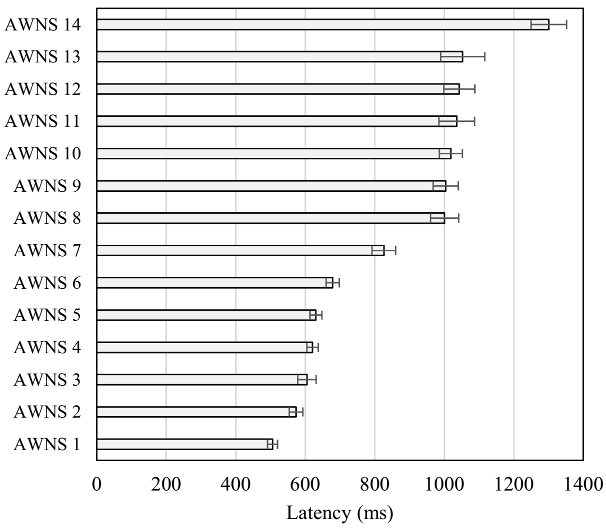

Figure 4: *Mean latencies of individual AWNS. Error bars indicate standard errors.*

latency of the AWS group was significantly shorter than that of the AWNS group (Wilcoxon rank sum test, $p < 0.05$).

Figures 3 and 4 show the individual mean shadowing latencies of the fifteen AWS and the fourteen AWNS, respectively. In the AWS group, 4 of the AWS showed latencies less than 500 ms, the remaining AWS had latencies ranging from around 500 ms to 900 ms. On the other hand, no one had the mean latency less than 500 ms in AWNS group: the mean latency of 7 AWNS ranged between 500 to 900 ms, while that of the other 7 AWNS were longer than 1000 ms. However, no one in the AWS group had a shadowing latency longer than 1000 ms.

3.2. Speech error comparison between AWS and AWNS

Figure 5 summarizes the speech errors in the two groups (median value: 3.34 vs. 0.96). The AWS as a group was significantly more error-prone than the AWNS (Wilcoxon rank sum test, $p < 0.05$), although there was a substantial overlap between the groups. 2 AWS were above the error rate range recorded by the AWNS (Figure 6).

In order to find out the relationship between the speech error rate and the latency, a scatter diagram is plotted in Figure 6. There was no significant correlation between speech errors and latencies as a whole ($\rho = -0.15$, n.s.). Either of the AWS and AWNS groups did not show a significant correlation between speech errors and latencies, either (AWS: $\rho = -0.30$, n.s.; AWNS: $\rho = 0.20$, n.s.). The results of the AWS group were subdivided into two groups at the latency of 500 ms, as shown with a solid vertical line in Figure 6. The boundary is just below the shortest mean latency of the individual AWNS participants. 4 AWS deviated below the range of the latencies of the AWNS, while a half of the AWNS had shadowing latencies above the range shown by the AWS. The 4 AWS whose shadowing latencies were shorter than 500 ms had a tendency of correlated latencies and speech errors ($\rho = -0.40$, $p = 0.10$). The eleven AWS whose shadowing latencies were more than 500 ms showed a significant negative correlation between the speech error rate and the latency ($\rho = -0.79$, $p < 0.05$). On the other hand, the AWNS group seems had a gap around 900 ms latency. However, both of the subgroups showed no significant correlation between speech errors and latencies (7 AWNS whose shadowing latencies ranged from 500 ms to 900 ms: $\rho =$

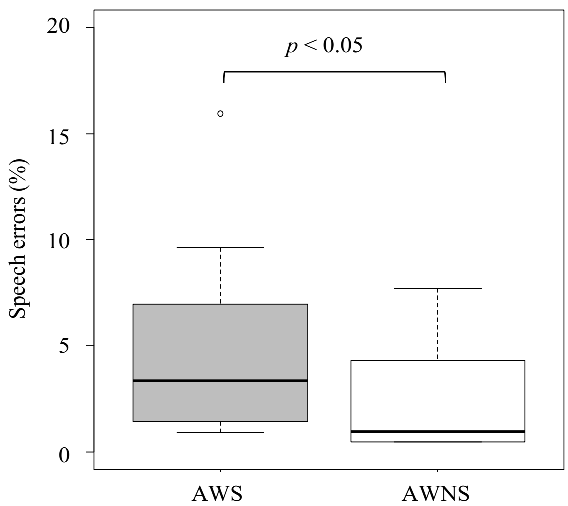

Figure 5: *Group comparison of speech errors.*

0.37, n.s.; another 7 AWNS whose shadowing latencies ranged more than 1000 ms: $\rho = 0.36$, n.s.).

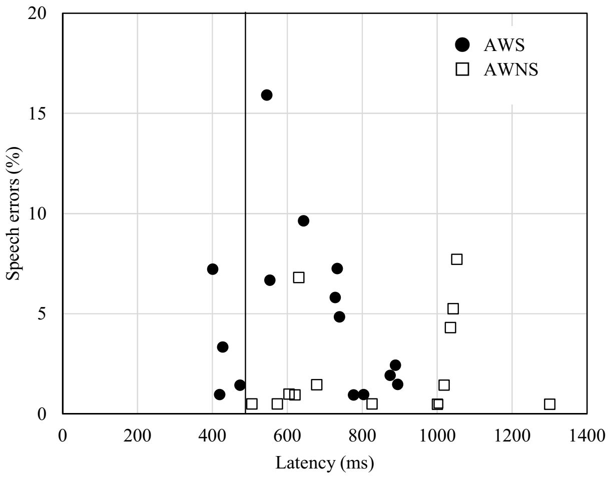

Figure 6: *The scatter plot of speech errors and latencies. The horizontal axis indicates the shadowing latency, and the vertical axis indicates the percentage of speech errors of individual subjects. The vertical line at 500 ms shows the lower limit of the shadowing latency in the AWNS.*

4. Discussion

The shadowing latencies in Japanese AWS and AWNS were investigated in the present study. The first finding of this study is that there were “close-shadowers” and “distant-shadowers” (shadowing latencies ranging from 506 ms to 1301 ms, with a possible gap at around 900 ms) in AWNS. This variation in the shadowing latency is consistent with those found in previous studies [5, 9, 10, and 26]. However, the latencies in the present study were longer than those previously reported for English [5]. This difference could be explained by the fact that the words with 3 or 4 syllables (morae) long are more common than single syllable words in Japanese [28, 29], whereas the highest frequency word lengths in English is 1 and 2 syllables [30, 31]. Except the difference of the absolute latencies, there were similar subdivisions into those with long and short latencies [5]. However, Healey (1987) [14] proposes three shadowing strategies according to the latency difference: “word-by-word”, “small phrases” and “large phrases”. In the current study, the shorter latency AWNS are likely to correspond to those who used a strategy of either word-by-word or small-phrases, or both. The reason why the two strategies did not separate from each other with the latencies may be due to the Japanese grammar that requires phrases to start with content words rather than post-positioned shorter function words, although it is possible that the few with the shortest latencies in the AWNS may have used the word-by-word strategy more often than the phrase-by-phrase strategy. In this subgroup, their latencies would have given the participants just enough time for a single phrase typically comprising several syllables, which however, seem to have been more than enough to shadow correctly, since the error rate did not correlate with the latency. The other subgroup whose latency was more than 1000 ms in the AWNS must have listened to a few phrases before they repeated, which corresponds to the large phrases strategy of Healey [14].

The second finding is that the AWS group had a significantly shorter shadowing latency than the AWNS group, which is not in line with a previous study [19]. The primary reason for this difference is likely due to the fact that Harbison & Porter (1989) [19] used vowel stimuli, with which listening and speech production were not exactly simultaneous, while in the current study meaningful longer materials (running speech) were used. Some of the AWS were shown to require as short as 500 ms for starting shadowing, with the error rates within the range found for the AWNS. The 500 ms is comparable to the duration of 1 word at the speech rate in the present model speech. Since the shadowing latency in Japanese for learners of Japanese as a L2 are 638-947 ms [11], which is longer than the shortest latencies of the AWNS in the present study, a possible lower skill level of Japanese speech due to stuttering does not fully account for the smaller latencies in the AWS than those in the AWNS. The short latency in the AWS may have been caused by some limitation of the working memory of AWS, because the latency-error trade off was observed in the subdivision of those who supposedly used the phrase-by-phrase strategy, judging from the latency range from 500 to 900 ms.

Thirdly, although the analyzed speech errors (word substitutions, word insertions, and word or part-word omissions) did not include dysfluencies specific to stuttering, the AWS made significantly more speech errors than the AWNS in the shadowing task. This is in line with Healey [14], in which AWS increased their mean number of word or sound omissions in shadowing than in a reading aloud task. Moreover, AWS tends to engage in stuttering coping behaviors such as word substitution, or omit particular words in their speech [32], probably they used this kind of coping behaviors during shadowing. Another previous study [5] showed that short-latency participants made more errors in their shadowing performance than long latency participants. If the negative correlation between the latency and error rate in the more than 500 ms latency subgroup of the AWS stands for a trade-off between these parameters, it suggests that they may have used their working memory almost to its limit during shadowing by the use of the small-phrase strategy.

5. Conclusions

The present study demonstrated significant differences in the shadowing latency and speech error rates in Japanese native AWS and AWNS in a speech shadowing task. The resultant latencies were longer than previously reported for English both in the AWNS and AWS groups, which most likely reflects the larger mean syllable numbers per word as well as per phrase of Japanese than of English. The AWS group had a significantly shorter latency than the AWNS. Besides, it was significantly more error-prone than the AWNS even after excluding dysfluencies specific to stuttering. The eleven AWS whose latency was more than 500 ms showed a significant negative correlation (trade off) between speech errors and latencies, which was not the case with the AWNS group. The results imply that the AWS may have hit the limit of their working memory capacity during shadowing while predominantly using a phrase-by-phrase strategy. Thus the shadowing paradigm brought new insights into speech processing and stuttering.

6. Acknowledgements

This research was supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research Grant Numbers 26770158 (first author), 23320083 (second author), and 60415362 (third author).

7. References

[1] Y. Mori, "Shadowing with oral reading: Effects of combined training on the improvement of Japanese EFL learner's prosody," *Language Education & Technology*. vol. 48, pp. 1-22, 2011.

[2] R.N. A and R. Hayashi, "Accuracy of Japanese pitch accent rises during and after shadowing training," *Proceedings of the 6th International Conference on Speech Prosody*, pp. 214-217, 2012.

[3] R.N. A, R. Hayashi, T. Kitamura, "Naturalness on Japanese pronunciation before and after shadowing training and prosody modified stimuli," *Proceedings of Speech and Language Technology in Education*, pp. 143-146, 2013.

[4] W.D. Marslen-Wilson, "Sentence perception as an interactive parallel process," *Science*, vol. 189, pp. 226-228, 1975.

[5] W.D. Marslen-Wilson, "Speech shadowing and speech comprehension," *Speech communication*, vol. 4, pp. 55-73, 1985.

[6] R.J. Porter and F. X. Castellanos, "Speech-production measures of speech perception: Rapid shadowing of VCV syllables," *Journal of the Acoustical Society of America*. vol. 67, no. 4, pp. 1349-1356, 1980.

[7] L. Scarbel, D. Beautemps, J. L. Schwartz, M. Sato, "The shadow of a doubt? Evidence for perceptuo-motor linkage during auditory and audiovisual close-shadowing," *Frontiers in Psychology*, vol. 5, 1-10, 2014.

[8] H. Mitterer, M. Ernestus, "The link between speech perception and production is phonological and abstract: evidence from the shadowing task," *Cognition*, vol. 109, pp. 168-173, 2008.

[9] T. Oki, "The role of latency for word recognition in shadowing," *The Japan Society of English Language Education*, vol. 21, pp. 51-60, 2010.

[10] S. Miyake, "Cognitive Processes in Phrase Shadowing: Focusing on Articulation Rate and shadowing latency," *The Japan Association of College English Teachers*, vol. 48, pp. 15-28, 2009.

[11] K. Kurata, N. Matsumi, "Nihongo Shadoingu-no ninchi mekanizumu-ni kansuru kiso kenkyu: bun-no onin imi syori-ni oyobosu gakusyusya-no kioku yoryo, bun-no syurui, bunmyaku-no eikyo [A basic study of the cognitive mechanism of shadowing in Japanese: the influence of memory span, sentence type, and context on phonological and semantic processing of sentences]," *Nihongo Kyoiku* (Journal of Japanese Language Teaching), vol. 147, pp. 37-51, 2010.

[12] E. Cherry, B. Sayers, "Experiments upon total inhibition of stammering by external control and some clinical results," *Journal of Psychomotor Research*, vol. 1, pp. 233-246, 1956.

[13] G. Andrews, P.M. Howie, M. Dozsa, BE. Guitar, "Stuttering: Speech pattern characteristics under fluency-inducing conditions," *Journal of Speech and Hearing Disorders*, vol. 25, no. 2, pp. 208-216, 1982.

[14] E.C. Healey, S.W. Howe, "Speech shadowing characteristics of stutters under diotic and dichotic condition," *Journal of Communication Disorders*, vol. 20, pp. 493-506, 1987.

[15] R.N. A, N. Sakai, K. Mori, "Tanki syadoingu kunren-no kitsuon-ni taisuru kooka [Short-term effects of speech shadowing training on stuttering]," *Onsei Gengo Igaku [The Japan journal of logopedics and phoniatrics]*, vol. 56, no. 4, 2015 (accepted).

[16] C. Van Riper, *The Treatment of Stuttering*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[17] O. Bloodstein, R.N. Bernstein, *A Handbook on Stuttering*. Delmar Learning, New York, 2008

[18] B. Guitar, *Stuttering: An Integrated Approach to Its Nature and Treatment*. Lippincott Williams & Wilkins, 2013.

[19] D. C. Harbison, R. J. Porter, "Shadowed and simple reaction times in stutterers and nonstutterers," *Journal of the Acoustical Society of America*. vol.86, no.4, 1277-1284, 1989.

[20] D. E. Cross, H. L. Luper, "Voice reaction time of stuttering and non-stuttering children and adults," *Journal of Fluency Disoders*, vol. 4, pp. 59-77, 1979.

[21] B. C. Waston, P. J. Alfonso, "Acoustic laryngeal reaction time: foreperiod and stuttering severity effects," *Haskins Laboratories: status report on speech research*, SR-71/72, pp. 261-280, 1982.

[22] E. Ozawa, Y. Hara, N. Suzuki, H. Moriyama, Y. Ohashi, *Kitsuon Kensaho [Stuttering Test]*, Gakuensha, Tokyo, 2013 [in Japanese].

[23] N. Sakai, J. Ogura, K. Mori, S. Y. Chu, Y. Sakata, "Nihongoban OASES-A-no Hyojunka - Genyuukai-niokeru Yobiteki Choosa - [Standardization of the Japanese version of the overall assessment of the speaker's experience of stuttering for adults (OASES-A) – a preliminary study with members of Japanese self-help groups-]," *Onsei Gengo Igaku [The Japan journal of logopedics and phoniatrics]*, vol. 56, 1-11, 2015.

[24] J.S. Yaruus and R.W. Quesal, *Overall assessment of the speaker's experience of stuttering (OASES), Pearson Assessments*, Bloomington, 2010.

[25] W. N. Patick and C.A. Fowler, "Shadowing latency and imitation: the effect of familiarity with the phonetic pattering of English," *Journal of Phonetics*, vol. 31, pp. 63-79, 2003

[26] S. Bultena, T. Dijkstra, J.G. V. Hell, "Switch cost modulations in bilingual sentence processing: evidence from shadowing," *Language, Cognition and Neuroscience*, pp. 1-20, 2014.

[27] P. Boeesma, D. Weenink, "Praat: doing phonetics by computer," Version 5.3.80. Available at: http://www.praat.org/, 2013 [accessed December 14, 2014].

[28] M. Sugito, *Word Accent in Japanese and English: What Are the Differences?* Hituzi Shobo, Tokyo, 2014.

[29] H. Kubozono, *Gokeisei-to Onin Koozoo [World Formation and Phonological Structure]*, Kurosio Publication, Tokyo, 1995.

[30] N. Umeda, "Linguistic rules for text-to-speech synthesis," *Proceedings IEEE*, vol. 64, pp. 443-445, 1976.

[31] N. Umeda, "Consonant duration in American English," *Journal of the Acoustical Society of America*. vol. 61, no. 3, pp. 846-858, 1977.

[32] M. Vanryckeghem, GJ. Brutten, N. Uddin, J.Van Borsel, "A comparative investigation of the speech-associated coping responses reported by adults who do and do not stutter," *Journal of Fluency Disorders*, vol. 29, no. 3, pp. 237-250, 2004.