A Corpus-based Analysis of Shadowing Speech: Case of L2 English by Japanese Learners

Shuju Shi[1,2], Nobuaki Minematsu[2], Yutaka Yamauchi[3]

[1]*University of Illinois - Urbana Champaign, USA,* [2]*The University of Tokyo, Japan,*

[3]*Tokyo International University, Japan,*

[1]*shujus2@illinois.edu*, [2]*mine@gavo.t.u-tokyo.ac.jp*, [3]*yyama@tiu.ac.jp*

Abstract

In this study we intend to investigate the typical phenomena in shadowing speech and work out a tentative scheme for shadowing speech labeling. Our aim is two-fold: a) to give useful feedback to students and teachers who are using shadowing as a way of language learning; b) to explore the possibility of automatic assessment and error detection of shadowing speech. We firstly labeled a shadowing corpus, which includes data from 10 male and 10 female students, where their speaking proficiency ranges from low, to intermediate to high level. Then we did statistical analysis on the labeled phenomena and our preliminary results are: 1) roughly defining ten prototypes of speech phenomena in shadowing, including substitution, insertion, omission, filled pause, repetition, grammatical errors(tense, voice, etc.), whispering, rearrangement of words' order, prosodic shadowing (shadowing without grasping the actual content of the speech) and combing parts from several words into one; 2) analyzing the subtypes within each prototype and finding that some types are gender/speaker dependent, and others are gender/speaker independent; 3) finding good correlation between our defined phenomena and shadowers' overall language proficiency.

Keywords: shadowing, Japanese L2 English, L2 acquisition, speech corpus, error analysis.

1. Introduction

Speech shadowing is a task where the subject is required to repeat speech as he/she hears it. Since shadowing includes processes of speaking, listening and comprehension of speech simultaneously, it was firstly introduced as a practicing strategy among simultaneous interpreters (see [1] for a review) and later language teachers also adopted this method. Recent decades have seen the effectiveness of shadowing in language learning [2-4], especially in Japan. Research in [2-3] showed shadowing can improve students' listening comprehension. The result in [2] also suggested that shadowing can enhance learners' phoneme perception ability. Research in [4] showed that shadowing can improve learners' intonation, fluency, word pronunciation and overall pronunciation. And comparison study suggested that shadowing could be more or at least no less effective than extensive reading, reading aloud and listening in improving speaker's corresponding language skills, that is reading comprehension, speaking, and listening comprehension [3,5-6].

The reason why shadowing could benefit language learning probably has its foundation in its processing mechanism. Other than simply repeating, shadowing has shown to involve complex production-perception interaction, automatic semantic and syntactic processing [7-8], and some people even performed sophisticated error correction during shadowing [9-10]. This, plus the fact that shadowing is a combined process of speaking, listening and comprehension, suggest that analytical results of shadowing speech can represent the speakers' overall language proficiency better than those of reading speech [11].

In our previous research, we realized automatic assessment of shadowing speech using the average of Goodness of Pronunciation (GOP) [12]. The result is promising with relative high correlation coefficient between automatic scores and speakers' TOEIC scores. But corrective feedback was not examined yet. For that, we have to know what kinds of phenomena tend to be observed in shadowing utterances. Few research have examined the production of shadowing speech, especially on corpus-level. In [13], to compare the performance of close (shadowing with a rather short latency) and distant (shadowing with a relatively long latency) shadowers, the authors analyzed the shadowing speech and defined three kinds of errors. They are as follow:

- ✧ Constructive errors: adding, substituting or creating new words to represent the presented words.
- ✧ Delivery errors: slurring hesitations, stuttering and unintelligible responses.
- ✧ Omission errors: omission of complete words.

The work offers us insights in investigating shadowing phenomena in our own task. In this study, we annotated 20 speakers of shadowing speech, defined 10 prototypes of typical shadowing phenomena and evaluated their role in reflecting speakers overall language proficiency.

2. Corpus description

Shadowing material we used is a passage about Fugu (puffer fish), which is a familiar topic to Japanese. It has 333 words in 21 sentences. The presented native speech was provided by an English teacher of native General American English speaker. The subjects (10 female, 10 male) are all university students and have never been exposed to the shadowing material before recording. They are allowed to practice shadowing with no reference to the text for 3-4 times before the final shadowing. Each student would take a simulated TOEIC test ahead of the recording. Their TOEIC scores are shown in Table 1..

Table 1: TOEIC scores of all the subjects.

Gender	TOEIC score
Female	955,940,895,825,601,592,581,308, 301,275
Male	990,990,968,625,436,395,367,289, 278,158

Classification of language proficiency level based on TOEIC score is shown in Table 2.

Table 2: Language proficiency level based on TOEIC.

level	TOEIC score
Low	0~399
Intermediate	400~699
High	700~990

3. Annotation and result

Annotation was done by the first author using Praat. In total, we defined ten prototypes of shadowing phenomena. Detailed explanation of each phenomenon, and analysis of annotation result are explained in detail in this section.

3.1. Typical phenomena in shadowing speech

Table 3 shows each phenomenon, its brief description, example, and labeling norm used in our research.

Table 3: Typical phenomena in shadowing speech.

| Name	Description and Labeling Norm
Substitution: 1)word-level 2)syllable-level	A(B)/A(<bcd>) means word A is substituted by word B or syllables <bcd>. e.g. The symptoms (sentence) e.g. expensive (<ikstin>)
Omission	A(-B) means the omission of word B e.g. had (-been) poisoned
Grammatical Errors	(sth.--sth.) defines errors that are related to tense and grammar and their combination. e.g.: Works → worked(tps--pt)
Insertion	(+B) means insertion of a word. e.g. (+the)
Repetition 1)syllable-level 2)word-level	Words are partly or fully repeated. e.g. over <+twi--> twice its e.g. very very(+1) expensive
Multi2One	A+B+..+N(=X) means a sequence of words are arranged as a cluster of syllables X. e.g. two hundred + dollars (=hudo)
Mimic	A(*) means word A is shadowed as some sound similar to the presented stimuli but the speaker actually didn't get the semantic meaning of the words.
Spoken Noise	Filled pause, e.g. <uh>, <en>, etc.
Non-spoken Noise	Noise other than spoken noise e.g. <microphone> ,<sniff>, etc.
Whispering	A(*whs) means word A is whispered because the speaker is not sure about what is presented in the stimuli.

3.2. Result of annotation

Figure 1 shows the overall result of the annotation. As can be seen, the first three most salient error types are omission, non-spoken noise and substitution. Students tend to keep silent while they cannot catch up with the presented stimuli and this explains why frequency of omission is the highest. The reason why frequency of non-spoken noise is so high perhaps lies in the fact that students are a little

bit nervous during shadowing. And also because shadowing itself costs high cognitive load, students do not realize that they are making unrelated noise. Whisper and multi2one are the least likely errors shown in Figure 1. In fact, the overall low frequency of these two types of errors also partly results from the fact that they are quite speaker-dependent. The frequency of multi2One for one female speaker is 38, accounting for 70% of overall frequency. And the situation for whisper is alike.
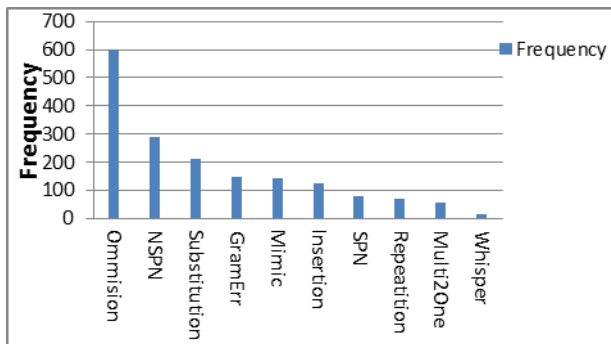
Figure 1: Overall result of labeling, where NSPN means non-spoken noise and SPN means spoken noise.

We further analyzed the distribution of each error type among different speaking proficiency levels. The result is shown in Figure 2. An overall tendency is that low level speakers tend to have more errors in each type except for non-spoken noise (NSPN). This suggests that the number of errors could serve as an indicator of the speakers' overall proficiency. One exception is the error type of non-spoken noise. As mentioned before, this type relates more to the state or in other words, speaking habits of the shadower, rather than his/her language proficiency.
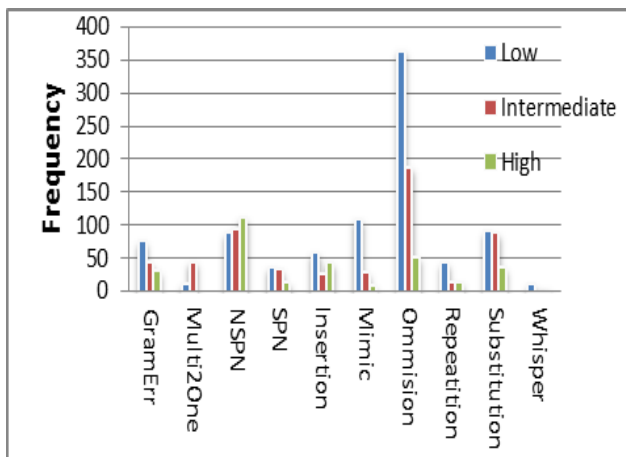
Figure 2: Result for different proficiency levels.

Figure 3 shows the female and male result separately. Generally speaking, compared with female speakers, male speakers tend to have less omission, substitution, insertion and more mimic, repetition, whisper.
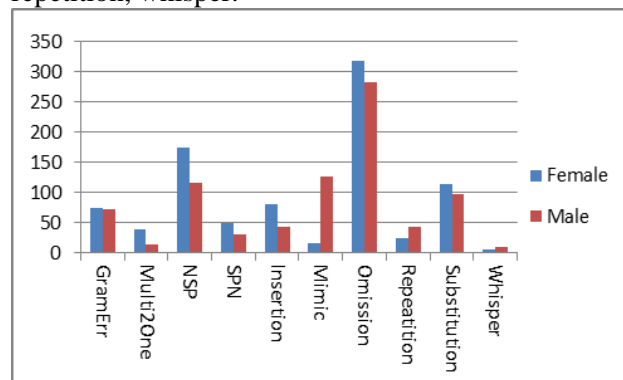
Figure 3: Result for male and female difference.

To see how well these error types are related to shadowers' overall language proficiency, correlation coefficients between the frequency of each error type and shadowers' TOEIC scores are calculated and given in Table 4. Error types that achieve statistically significant correlation with TOEIC scores are omission, grammatical errors, mimic and repetition. And the correlation between frequency omission and TOEIC score is rather high (-0.802). This suggests that in order to realize automatic assessment and corrective feedback of shadowing speech, we need to pay close attention to these types of errors.

Table 4: Correlation Coefficients between frequency of error types and TOEIC scores.

Error Type	TOEIC scores
Omission	-.802**
GrammErr	-.546*
Mimic	-.499*
Repetition	-.474*
Substitution	-.430
Insertion	-.223
SPN	-.200
Whisper	-.132
Multi2One	-.092
NSP	-.081

** denotes that significant level is 0.01.
* denotes that significant level is 0.05.

4. Discussion

In this section, we discussed some specific findings in the annotation process, pointed out limitations about this research and gave possible directions for future research.

4.1. Same error type, different strategy

Even though learners of three different proficiency levels share the same error types in our current labeling norm, the underlying mechanism is quite different. High level learners tend to maintain syntactic correctness and semantic connection. For example, in the utterance "… their hand at preparing the fish themselves", one high-level speaker mis-shadowed the word 'at' as to, meanwhile she changed 'preparing' to 'prepare'. This tendency of keeping syntactic and semantic correctiveness is also found in [13], which analyzes shadowing in native speakers. On the other hand, errors by low level learners usually reflect their inability to catch what's in the presented stimuli or to repeat what they got correctly.

4.2. Female and male difference

The shadowing strategy of female and male are quite different when they encounter something they cannot comprehend. When Female learners missed the presented stimuli, they tended to keep silent or uttered some filled pause, while male learners would follow the stimuli and uttered some non-meaningful but prosodic similar sounds.

4.3. Speaker-dependency or/and material-dependency

In our study, some error types such as multi2One and whisper tended to be speaker-dependent. For this moment, we could not be sure whether these error types are truly speaker-dependent or it is because the number of subjects is not big enough in our dataset.

Also, shadowers' performance can be highly influenced by the difficulty degree of the presented stimuli. In this research, we used a carefully selected intermediate level of passage. We would like to further investigate how learners behave in easier/ more difficult situations.

5. Conclusions

In this study, to investigate typical phenomena in shadowing speech, we manually annotated English speech shadowed by 20 Japanese learners. In the annotation process, we defined ten prototypes of typical error types and further analyzed the annotation results. What we found are: 1) distribution of error types on different language proficiency levels and the correlation coefficient results showed that frequency of error types we defined could serve as good indicator of shadowers' overall language proficiency; 2) some error types are gender-dependent or even speaker-dependent; 3) like in native shadowing, shadowing in a second language also includes syntactic and semantic processing of the speech and this is even true among high level learners.

7. References

[1] Lambert S. Shadowin,. 1992. *Meta: Journal des traducteursMeta:/Translators' Journal*, 37(2): 263-273.

[2] Y. Hamada, 2014.The effectiveness of pre-and post-shadowing in improving listening comprehension skills. *The Language Teacher*, *38*(1), 3-10.

[3] Y. Hamada, 2015. Shadowing: Who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research*.

[4] K. T. Hsieh, D. A. Dong, & L. Y. Wang, 2013. A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics*, *11*(2), 43-66.

[5] T. Nakanishi,, & A. Ueda, 2011. Extensive reading and the effect of shadowing. *Reading in a Foreign Language*, *23*(1), 1.

[6] A. Kuramoto, O. Shiki , H. Nishida, & H. Ito, 2007. Seeking for effective instructions for reading: The impact of shadowing, text-presented shadowing, and reading-aloud tasks. *LET Kansai Chapter Collected Papers*, *11*, 13-28.

[7] H. Mitterer, & M. Ernestus, 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task.*Cognition*, *109*(1), 168-173.

[8] P. W. Carey, 1971. Verbal retention after shadowing and after listening.*Perception & Psychophysics*, *9*(1), 79-83.

[9] W. D. Marslen-Wilson,. 1975. Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226-228.

[10] W. D. Marslen-Wilson, 1973. Linguistic structure and speech shadowing at very short latencies. *Nature*.

[11] D. Luo, N. Minematsu, Y. Yamauchi, & K. Hirose, 2009. Analysis and comparison of automatic language proficiency assessment between shadowed sentences and read sentences. In *SLaTE* (pp. 37-40).

[12] D. Luo, N. Minematsu, Y. Yamauchi, & K. Hirose, *2008.* Automatic assessment of language proficiency through shadowing. *International Symposium on Chinese Spoken Language Processing, ISCSLP'08. 6th International Symposium on* (pp. 1-4).

[13] W. D. Marslen-Wilson, 1985.Speech shadowing and speech comprehension. *Speech Communication*, *4*(1-3), 55-73.