

Reading Aloud Performance and Listening Ability in an L2: The Case of College-Level Japanese EFL Users

Shigeo Kato¹, Kento Tanaka²

¹Faculty of Education, Niigata University, Niigata, Japan

²Tsubame Secondary School, Niigata, Japan

Email: skato@ed.niigata-u.ac.jp

Received 16 March 2015; accepted 15 April 2015; published 20 April 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Recent advance in neuropsychology has evidenced a facilitative role of sensorimotor activity for the development in L2 speech perception. The study attempted to examine the relationship between reading aloud (RA) performance, grammatical knowledge and listening ability with 31 college-level Japanese EFL users. The result demonstrated highly significant correlations between all the variables, and the subsequent multiple regression analysis also indicated RA significantly accounting for listening. Supplementary analyses dividing the participants by listening ability demonstrated that while significant correlation was maintained between L2 knowledge and listening with less-proficient listeners, it disappeared with proficient listeners; in contrast, significant correlation between RA and listening performance was maintained in both groups, indicating that production accuracy/fluency still played an important role in advanced L2 listening.

Keywords

Reading Aloud, Listening Ability, Sensorimotor Activity, English as a Foreign Language

1. Introduction

The *motor theory of speech perception* by Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967), which suggests that speech perception depends on access to the speech motor system, and the later work by Liberman and Mattingly (1985), which demonstrates that speech perception is specifically facilitated by coordination between the *perceived gestures* of the speaker's vocal tract and matching *intended gestures* on the part of the listener (p. 3), has not appeared to be greatly influential in research into the L2 listening process. However,

recent developments in neuroscience, in particular the discovery of the *mirror neuron system*, have inspired the reconsideration of the implications of the motor theory for the mechanism of speech perception. This system was first reported by a research team led by Giacomo Rizzolatti, who, while studying the activation of the premotor cortex in macaque monkeys in the perception of hand movements, unexpectedly found a system of neurons firing both when the monkeys performed an action themselves and when they watched a researcher performing the same action (e.g., di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). This has provided evidence that “in addition to action observation eliciting concurrent performance of that same action, the performance of an action influences the concurrent perception of that action” (Oberman, Winkielman, & Ramachandran, 2007: p. 167). Further, mounting empirical evidence from speech-related brain imaging studies has indicated the facilitative involvement of specific motor circuits during speech perception (e.g., Casserly & Pisoni, 2010; Gandour et al., 2007; Iacoboni, 2008; Skipper, Nusbaum, & Small, 2005)—an insight that is leading to a critical paradigm shift in our understanding of the L2 listening process, moving us from the longstanding view of the perception-to-production sequence (e.g., Derwing, Thomson, Foote, & Munro, 2012; Thomson, 2012), that sees perception and production as separate modules, to one that sees these two processes interacting and facilitating each other (e.g., Casserly & Pisoni, 2010; Cogan, Thesen, Carlson, Doyle, Devinsky, & Pesaran, 2014).

However, the contribution of production performance to listening ability has yet to be examined in an L2 context. The present study thus attempts to directly investigate the impact of L2 production performance, as measured by the accuracy and fluency of reading aloud by Japanese university-level EFL learners, on their L2 listening proficiency, with L2 grammatical knowledge as an additional explanatory variable.

2. Literature Review

2.1. Implications of Neuroimaging Studies

The growing number of neuroimaging studies focusing on the activation of brain regions while listening can provide evidence with a critical impact for our understanding of speech perception. To be more specific, the discovery of *sensorimotor neurons* that are active during both action execution and corresponding perception (e.g., Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008) can yield robust empirical support for the idea that the ability to articulate the target language’s sound system has a positive impact on the acquisition of L2 perception ability. Such studies have indicated that the main brain areas that have been traditionally linked to speech production (i.e., the pars opercularis and the triangularis of the inferior frontal gyrus \approx BA 44/45 \approx Broca’s area) are also activated in speech perception (for an overview of these findings, see Price, 2012). For example, Londei et al. (2010) used functional magnetic resonance imaging (fMRI) to investigate the brain activity of participants passively listening to words, pseudo-words and reversed-played words in their L1. The results suggested that the reproducibility of an incoming speech stimulus is a critical feature in the regulation of the speech perception network. Furthermore, “learning and development of speech production might shape the mapping between sensory and motor maps that later might become useful in predicting and generating hypotheses on the incoming information” (p. 578). Indeed, Meister, Wilson, Deblieck, Wu, and Iacoboni (2007) demonstrate that disruption of the premotor cortex impairs speech perception and that activity in motor areas is causally linked to speech perception as well.

L2 sound perception, Wilson and Iacoboni (2006) used fMRI to investigate neural responses to familiar (native) and unfamiliar (non-native) phonemes among adult monolingual English speakers. Their results indicated that the motor areas play an important role in distinguishing native and non-native phonemes, and more interestingly, that the motor system engages in repeated attempts to perceive heard non-native phonemes, leading to greater, and hence likely more costly, motor activity (ibid., p. 322). Callan, Jones, Callan, and Akahane-Yamada (2004) used event-related fMRI to investigate brain activity related to perception by native Japanese speakers who were late-onset English learners (that is, who began learning it after childhood) of the English /r-/l/ phonetic contrast, which is non-existent in the Japanese sound system. Their findings suggest that greater differential activity in brain regions involved in speech production planning (specifically, Broca’s area and the sensorimotor cortex) was evidenced for the perceptual identification of /r/ and /l/ relative to that of vowels for Japanese L1 participants. Gandour et al. (2007), with a sample of Chinese-English bilinguals with late English onset and upper-intermediate ability (TOEFL score of 600 or more), provided similar findings for sentence-level prosodic features in English: an essential role played by the brain regions responsible for phonological processing and

speech motor planning and execution in the performance of an auditory sentence-focus judgment task (specifically, sentence-initial vs. sentence-final position of contrastive stress). Similar findings were reported in Wang, Sereno, Jongman, and Hirsch (2003), where the learning of Mandarin lexical tone by native English speakers was associated with the emergence of new activity in Broca's area. All these findings provide robust support for the assertion that motor areas play a critical role in speech perception.

Equally important, these studies provide evidence for developmental changes in cortical representation as a function of language proficiency. To be more specific, greater activity in the relevant brain regions was evident in Japanese L1 speakers than in English L1 speakers during perceptual identification of /r/ and /l/ (Callan et al., 2004), and also when participants performed a target task in an L2 (English) as compared to an L1 (Chinese) (Gandour et al., 2007). A point of note here is that the participants in both of these studies were late-onset bilinguals. Furthermore, in Gandour et al. (2007), differential activation in relevant brain regions was evidenced for participants with lower L2 proficiency (see also Chee, Hon, Lee, & Soon, 2001; Hasegawa, Carpenter, & Just, 2002; and Xue, Dong, Jin, & Chen, 2004, for similar findings). These findings have important implications for our understanding of the effect of accurate (subvocal) articulation ability for L2 phonological features on L2 listening ability, in particular among less proficient listeners.

2.2. Research into the L2 Listening Process

Second language (L2) listening comprehension is a multifaceted, multilayered skill involving various component operations, including but not necessarily limited to *phoneme-level perception*, *word recognition*, *lexical access*, *morphological and syntactic processing*, *activation of prior knowledge*, and *utilization of contextual information*, all interacting to produce a final representation (e.g., Field, 2008; Rost, 2002; Vandergrift, 2011). There have not been many studies focusing on how and to what relative degrees these factors explain L2 listening performance. Of the few, one pivotal work is Vandergrift (2006), whose results for a group of adolescent (14 - 15-year-old) English L1 speakers learning French in an L2 setting in Canada indicated that L1 listening ability and L2 proficiency together accounted for about 39% of L2 listening ability, the former explaining about 14% and the latter about 25%. Another study examining the contribution of L2 proficiency to L2 listening ability is Mecarty (2000), whose results from multiple regression analysis of college-level L1 speakers of English learning Spanish indicated that lexical knowledge but not grammatical knowledge significantly predicts listening comprehension, explaining 14% of total variance.

In general, the broad range of studies examining the impact of L1 literacy and L2 proficiency on L2 reading have had similar outcomes to those above considering L2 listening: while both L1 reading ability and L2 proficiency play important roles in successful L2 reading performance, the latter makes a greater contribution among beginning readers, and the predictive value of L1 literacy becomes greater the more advanced the learner's L2 reading level becomes (e.g., Bossers, 1991). In terms of the level of importance of specific components of reading ability for reading ability as a whole, research findings rooted in the concepts of *automaticity* (e.g., Segalowitz, 2003) and *working memory capacity* (e.g., Baddeley, 1986; 2000) suggest that once readers have reached a certain level of ability in *bottom-up processing*, encompassing for example orthographic, phonological, and lexical processing, more attentional resources can be allocated to *top-down processing* (e.g., Crosson & Lesaux, 2010; Farnia & Geva, 2011). The exact level of bottom-up processing ability needed is a function of L2 reading proficiency (e.g., Carrell, 1991), task type (e.g., Bernhardt & Kamil, 1995), task complexity (e.g., Taillefer, 1996), and L1 - L2 linguistic relationship (e.g., Koda & Zehler, 2008).

As processes of *receptive comprehension*, listening and reading share a significant number of component processes, such as *decoding*, *parsing*, and *meaning-building* (e.g., Mecarty, 2000; Vandergrift, 2006). Indeed, research findings from studies in L2 listening suggest that the level of contribution from basic L2 processing skills, such as efficiency of phoneme-level recognition and word segmentation, is much greater in beginning listeners than in their advanced counterparts (e.g., French, 2003) and that this is particularly the case in listening to concatenated speech (e.g., Goh, 2000; Graham, 2006). A range of research findings on the use of *listening strategies* has indicated that while skilled listeners make use of a wider range of both cognitive and metacognitive strategies, in particular of strategies to self-regulate their listening process, such as *elaborating questioning* and *monitoring comprehension*, and have higher self-awareness of their own listening problems (e.g., Goh, 2000), less skilled counterparts are likely to rely on bottom-up aspects of processing such as word-by-word lexical access (e.g., Graham, 2006; Osada, 2001) and on-line translation (e.g., Vandergrift, 2003). Findings from Gra-

ham (2006), with a sample of intermediate-level, adolescent (aged 16 - 18) English L1 speakers learning French reflecting on their own listening problems, indeed suggest that their listening difficulties appeared to stem from a lack of awareness of pronunciation and intonation features that are crucial for accurate French perception. Goh (2000) reported that less proficient listeners had remarkably greater difficulty in phoneme- and word-level recognition compared to more proficient listeners. These insights and those cited above collectively imply strongly that with listening as with reading, the less effortful bottom-up processing becomes, the greater the amount of attentional resources that can be allocated to top-down processing (e.g., Field, 2008).

2.3. L2 Listening Pedagogy

A range of practical teaching techniques have been proposed to develop lower-level processing efficiency in L2 listening, including *focused analysis* of the target script (Goh, 2002), the *word-spotting task* (Al-Jasser, 2008), *dictation* (e.g., Kiany & Shiramiry, 2002), *dictogloss* (e.g., Wilson, 2003), *exposure to “i-1” level passage* accompanying the script of the materials (Hulstijn, 2001), and various other remedial exercises. However, most of these techniques are intended to work only with repeated exposure to aural text, sometimes with varied speed control, and ultimately they mainly aim to provide learners with opportunities to “accumulate and categorize acoustic, phonemic, syllabic, morphological and lexical information” (Hulstijn, 2003: p. 422). However, the neuroimaging studies cited above provide a robust empirical basis for the effectiveness of instructional techniques aimed at developing articulatory accuracy/fluency to establish reliable *productive*, as opposed to *receptive*, phonology in the L2 (see Walter, 2008, for a similar argument), in particular for beginning L2 users whose L2 pronunciation ability is not yet well developed.

Despite the view of several reading researchers that reading aloud activities have not always been favorably considered in L2 classrooms, possibly because of “the misuse of the technique [of reading out loud] around the class” (Nation, 2009: p. 66), the usefulness of this approach for the development of lower-level processing efficiency has been widely confirmed in L2 reading research (e.g., Birch, 2007; Janzen, 2007; Gibson, 2008). In particular, reading aloud helps L2 learners establish accurate phonological representations (e.g., Gibson, 2008), encouraging their awareness of not only segmental but also suprasegmental features, such as rhythm, stress and intonation, by using connected texts rather than decontextualized vocabulary items (e.g., Kato, 2012). There has been a range of research in this regard among Japanese EFL learners, which has generally found that reading aloud significantly improves silent reading rate (Suzuki, 1998), reading performance (Miyasako, 2008), and reproduction of key words and phrases (Shichino, 2006). Miyasako (2008), for instance, investigated the contribution among upper-secondary level Japanese EFL users of six weeks of reading aloud practice for L2 reading performance; it was found that it significantly improved phonological decoding and reading comprehension performance, and that this practice effect was more pronounced with less proficient readers.

A range of relevant studies has been conducted using Elicited Imitation Test (EIT) or Elicited Oral Response (EOR) testing, with the aim of exploring the validity and reliability of these measurements as effective assessment tools particularly for L2 oral proficiency (e.g., Vinther, 2002; Cox & Davies, 2012; Wu & Ortega, 2013). These tests require the participants to listen and then repeat as exactly as possible, often after a brief interval of silence, a number of sentences with either a constant or increasing length; they show a significant effect on L2 speaking ability. Cox and Davies (2012) examined effects on speaking proficiency of several factors: EOR scores calculated by an automatic speech recognition (ASR) system and scores on listening, reading, writing, and grammar as measured by other tests, among English L2 users with various L1 backgrounds. Although their results indicated significant correlations for all the variables, the highest correlation was found between EOR performance and the listening test. However, their scores were based on the number of words repeated correctly enough to be recognized by the ASR system, not on the actual quality of productions compared with the stimuli. Nevertheless, their conclusion that “listening ability is an important component of both conversational speaking and the ability to process and repeat phrases in a second language” (p. 614) is supported; as mentioned, the reverse is also true: repetition ability can be a pre- or co-requisite of listening ability.

3. Methodology of Study

3.1. Participants

The participants were 33 adult native speakers of Japanese (17 women and 16 men) enrolled in undergraduate

and graduate courses in the faculty of education at a university in Japan. The mean age of the participants was 21.7 years old, with a range from 20 to 24.

3.2. Procedure

The tests were carried out in two separate sessions. The listening comprehension test and grammaticality judgment test (task) were conducted in the first session; all participants simultaneously took the tests together in a dedicated language laboratory at the university. After an interval of 4 to 7 days, the reading aloud task was conducted individually in a dedicated room at the university, as a second session. The first session lasted about 30 minutes and the second session about 15 minutes for each participant.

- Listening Comprehension Test

The material used for this test was adapted from the listening section of an exercise book for the TOEIC Test (Iwamura & Smillie, 2007). The test comprised three parts, each of which contained 15 multiple-choice questions (for a total of 45 questions). The duration of this test was approximately 25 minutes, including time for instructions. The number of correct answers (out of 45) was used as data in the subsequent analyses.

- Grammaticality Judgment Task

For this test, 45 items were developed by the authors. The word count of these sentences ranged from 4 to 13, with a mixture of simple, compound, and complex sentences. The vocabulary and grammatical structures were all covered in the national English curriculum at the lower secondary level (i.e., junior high school) or the first year of the upper secondary (high school) level (respectively corresponding to 7th-9th grade and 10th grade in North American schools). The goal was to provide test sentences of varying degrees of difficulty and so to ensure sufficient discriminative power. Grammatical error types tested included number agreement, as in *Kevin has three blue shirt*; declension, as in *Carol is cooking dinner for hers family*; inflection, as in *He wore his new hat yesterday* and *The people looked surprising when they heard her song*; word order, as in *I do not know what do you mean*; word choice, as in *This is a book what I bought in the United States*; and tense, as in *They have known each other for ten years when they got married*. Half of the target sentences were grammatically correct.

All the test sentences were presented as a list on a sheet of A4-sized paper; participants were requested to read each sentence silently and answer whether it was grammatically correct (without giving further details on the nature of the error if any). They were requested to start from the beginning and to answer as quickly and accurately as possible within a time limit of 5 minutes for the full task. Further, they were instructed to skip sentences only when they did not have a clue about the answer, in order to avoid random guessing, and otherwise to proceed in order down the list.

- Reading Aloud Task

The material was taken from *New Edition Unicorn English Course 1* (Ichitani, 2011), which an English textbook adopted in the national curriculum for tenth grade. The test text was 127 words long; its Flesch-Kincaid Grade level was 8.04 and its Flesch Reading Ease index was 62.4 (Kincaid, Fishburn, Rogers, & Chissom, 1975). The topic of the text was Natsume Soseki (1867-1916), a major Japanese novelist and also a researcher in English literature. The text briefly presents an episode from Natsume's widely familiar (in Japan) time on a research fellowship in the United Kingdom from 1900 to 1902. Given this text selection, it is reasonably unlikely that the language level or topic of the text led to significant variation in comprehension accuracy among the participants, since all of them had completed the national secondary curriculum for the subjects of English (covering the vocabulary and grammar used in the text) and Japanese (covering the content). The text is presented in Appendix.

The overall testing procedure and assessment scheme were adopted from Shimizu (2009). First, the participants were requested to read the text silently for 1 minute, and then to read it aloud at a normal speed. They were asked to perform an expressive reading aloud that could make the content comprehensible to an imagined audience. The rating was conducted in terms of the three criteria mentioned above: *segmental features*, *supra segmental features*, and *phrasing*. The criterion of segmental features focused on the correctness of individual phonemes. The criterion of supra segmental features focused on the sound features such as intonation, rhythm and sentence-level stress. The criterion of phrasing focused on the correctness of segmenting, or chunking, within each sentence in accordance with meaning. The rating scale for these criteria is presented in Table 1, adapted from Shimizu (2009, p. 183). A full score for each criterion was 5 points, and thus a total of 15 points were possible for each reading aloud performance.

Table 1. Rating scale for the reading aloud test (adapted from Shimizu, 2009: p. 183).

	Description of each criterion
Segmental features (1 - 5 points)	5 = No incorrect pronunciation of segmental features. 4 = A few incorrect pronunciation of segmental features. 3 = Several incorrect pronunciation of segmental features. 2 = Many incorrect pronunciation of segmental features. 1 = Too many incorrect pronunciation of segmental features.
Supra segmental features (1 - 5 points)	5 = No incorrect pronunciation of supra segmental features. 4 = A few incorrect pronunciation of supra segmental features. 3 = Several incorrect pronunciation of supra segmental features. 2 = Many incorrect pronunciation of supra segmental features. 1 = Too many incorrect pronunciation of supra segmental features.
Phrasing (1 - 5 points)	5 = No incorrect phrasing. 4 = A few incorrect phrasing. 3 = Several incorrect phrasing. 2 = Many incorrect phrasing. 1 = Too many incorrect phrasing.

All performances were audio-recorded. Each was scored by two raters, one a native speaker of English with an MA degree in education and the other a native speaker of Japanese with a PhD in applied linguistics; each rater had more than 10 years' experience teaching English language teaching (ELT) courses at a Japanese university. The findings of Shimizu (2009) suggest that the use of two experienced raters with a rating scale featuring three criteria provides the best reliability and cost efficiency for the measurement of reading aloud test results (p. 189). The raters in the current study first scored the recorded performances of each participant individually. They then got together to check all the performances to see if there were any differences between their scores of 2 points or more in any criterion, and held further discussion to reduce these gaps to 1 point or less. The final scores used for the subsequent analyses were calculated by adding the scores of each rater, and hence ranged from 6 to 30.

To sum up, the present study attempts to examine the explanatory power of two independent variables, namely, *reading aloud performance* and *L2 grammatical knowledge*, on *L2 listening ability* as a dependent variable. Following Shimizu (2009), reading aloud performance was measured in terms of three categories: individual sound-production accuracy (henceforth, *segmental features*), supra segmental sound-production fluency (henceforth, *supra segmental features*), and chunking (or phrasing) accuracy (henceforth, *phrasing*). Grammatical knowledge was measured on the basis of a grammaticality judgment task. The specific research questions for this study were set as follows:

1) *How are L2 reading aloud performance, as measured in terms of three components—the production performance of segmental features, supra segmental features, and phrasing—and L2 grammatical knowledge, as measured by the grammaticality judgment task, related to L2 listening performance?*

2) *Do the relationships found in the first research question, if any, reveal any change as a function of L2 listening proficiency?*

4. Results

Descriptive statistics for all measures for the participants as a whole appear in Table 2. Figures 1-3 present boxplots of the listening comprehension, reading aloud, and grammaticality judgment scores respectively. Boxplots are used instead of barplots on the basis of the insight of Larson-Hall and Herrington (2010) that the former graphics provide richer information about the dispersion and skewness of scores and the existence of *outliers* (p. 370). Bold lines in the boxes indicate medians, and the top and bottom end of each plot (Larson-Hall and Herrington call them “whiskers”) show the minimum and maximum scores of the distribution. The ends of the box show the 25th and 75th percentiles of scores, or the *interquartile range* (IQR); beyond 1.5 times the IQR above or below the box are outliers (Larson-Hall & Herrington, 2010: p. 371). As shown in the boxplot for the reading aloud task, there are two outliers on the reading aloud measure, one of whom is also an outlier on the listening comprehension test and the other on the grammaticality judgment test. Therefore, all data for these two participants was excluded from the subsequent analyses.

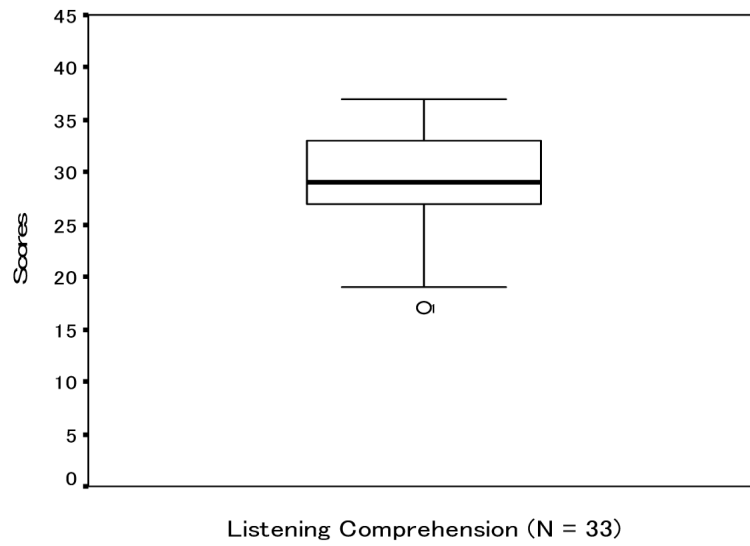


Figure 1. Boxplot for listening comprehension measure.

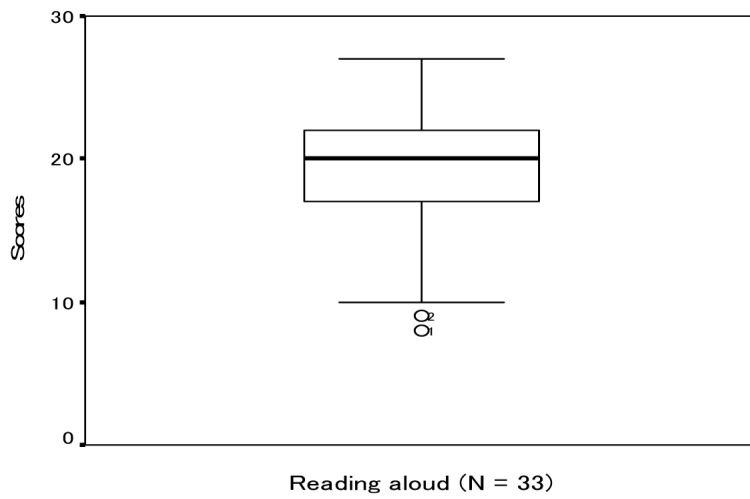


Figure 2. Boxplot for reading aloud measure.

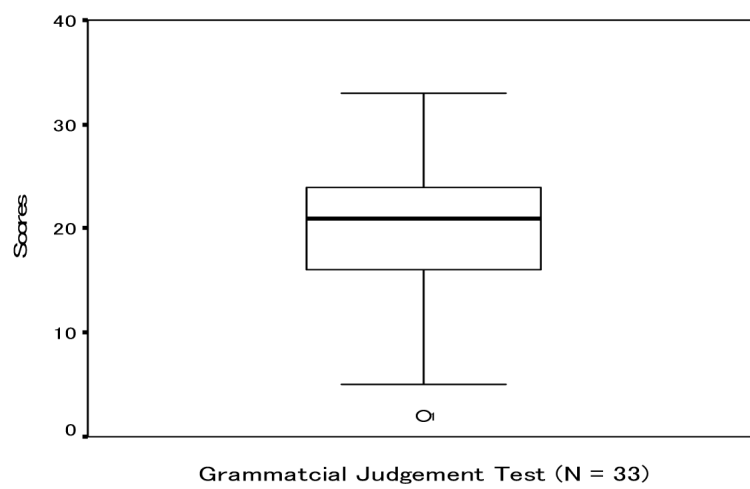


Figure 3. Boxplot for grammatical judgement measure.

Table 2. Mean values and standard deviations for listening comprehension, reading aloud (total, pronunciation, segmentation, and fluency), and grammaticality judgment measures for the participants as a whole. (N = 31)

	Score range	M	SD
Listening	0 - 45	29.2	4.67
Reading aloud total	6 - 30	20.1	3.96
Segmental features	2 - 10	6.32	1.45
Supra segmental features	2 - 10	6.65	1.52
Phrasing	2 - 10	7.16	1.34
Grammaticality judgment	0 - 45	20.5	6.75

Figure 4 shows multiple scatter plots between all the variables, with a loess curve. Since the sample size of the current study is under 50, a Shapiro-Wilk and not a Kolmogorov-Smirnov goodness-of-fit test was conducted to confirm the normality of the distribution. The results revealed normal distribution for listening, segmental features, and grammaticality; however, phrasing scores did not indicate a normal distribution (Statistic = .910, $df = 31$, $p < .05$) and the scores for supra segmental features were marginal (Statistic = .936, $df = 31$, sig. = .082). As seen in the corresponding scatterplots in **Figure 4**, this indicated that these two measures were negatively skewed, leading the null hypothesis (that the sample is normally distributed) to be rejected (e.g., [Bachman, 2004](#)). To solve this problem, log-transformation was performed on the scores from all three components of the reading aloud measure, including the scores for segmental features along with supra segmental and phrasing scores in order to obtain consistent data across the three components. Only transformed values were used in the subsequent analyses (e.g., [LaFrance & Gottardo, 2005](#)).

After the data transformation, Pearson product-moment correlation coefficients were conducted to examine interrelationships between all the measures, including components of total reading aloud scores. **Table 3** shows the results.

Next, hierarchical regression analysis was conducted to examine the predictive value of the four explanatory variables (L2 grammaticality judgment scores and the three components of reading aloud performance) for L2 listening comprehension performance as a dependent variable. In the first model (Model 1), grammaticality scores were entered first to partial out the shared variance of this variable. The order of entry of the three components in the following three models (Models 2 to 4) was based on the hierarchical level of the phonological unit in focus; that is, the component of segmental features, focusing on the individual phonemes, is at the lowest level; next comes the component of supra segmental features, covering rhythm, stress, and intonation; and, finally, the phrasing component, covering segmentation/chunking and sentential stress. Cook's distance and Mahalanobis distance values were checked to see if there were any influential outliers in this particular analysis, but the results suggested that there were not: the mean value for Cook's distance was .075 and that for Mahalanobis distance was 3.871 (see, e.g., [Field, 2005](#)). Variance Inflation Factor (VIF) was also checked to see if there was a possibility of multicollinearity, but in spite of the high level of intercorrelations among the explanatory variables, the result yielded did not indicate any concerns for the problem at hand: the highest value was 4.186, for fluency, from the fourth model of the regression analysis (cf. [Larson-Hall, 2010](#); [Heiberger & Holland, 2004](#)). **Table 4** shows the results of the analysis. The first regression model found L2 grammatical knowledge to be a significant predictor of L2 listening ability, accounting for 40% of the variance: $\beta = .632$, $t(29) = 4.396$, $p < .001$. The second model showed that while the predictive value of L2 grammatical knowledge turned out to be marginally significant, $\beta = .191$, $t(2, 28) = 1.798$, sig. = .083, the component of segmental features accounted for an additional 39% of the variance in L2 listening, $\beta = .764$, $t(2, 28) = 7.187$, $p < .001$. The third model indicated that the component of supra segmental features accounted for an additional 8% of the variance, $\beta = .502$, $t(3, 27) = 4.076$, $p < .001$, and the fourth model showed that the phrasing component accounted for an additional 5% of the variance: $\beta = .429$, $t(4, 26) = 0.3.688$, $p < .01$.

In order to investigate the second research question—how the relationship between the explanatory variables *reading aloud performance* (three components plus their total scores) and *grammaticality judgment scores* and the dependent variable *L2 listening ability* changes as a result of *proficiency in L2 listening*—the participants were divided into two groups (*proficient* and *less proficient* listeners) on the basis of their scores on the listening

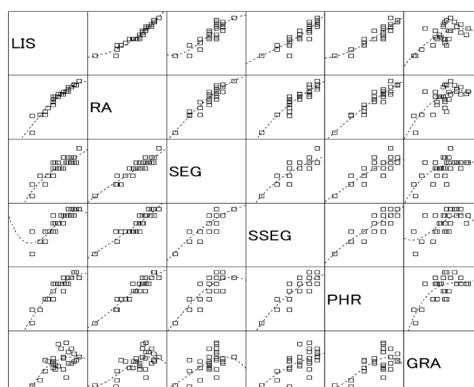


Figure 4. Multiple scatter plots between all the variables. Note: LIS = listening comprehension; RA = reading aloud; SEG = segmental features; SSEG = supra-segmental features; PHR = phrasing; GRA = grammaticality judgment.

Table 3. Bivariate correlations with reading aloud (total, pronunciation, segmentation, and fluency) and grammaticality against listening comprehension (N = 31).

	Listening		Reading aloud			
	Total	Seg	Sseg	Phr	GRA	
Reading aloud total	.96**					
Segmental features	.88**	.93**				
Suprasegmental features	.92**	.93**	.80**			
Phrasing	.89**	.94**	.80**	.84**		
Grammaticality judgment	.63**	.65**	.58**	.63**	.62**	

Note: **Correlation is significant at the .01 level (two-tailed). Seg = segmental features, Sseg = suprasegmental features, Phr = phrasing.

Table 4. Results of hierarchical regression analysis examining predictors of L2 listening performance.

Model	R	R ²	ΔR^2	Grammar (β)	Seg (β)	Sseg (β)	Phr (β)
1	.632a	.400	.400	.632**			
2	.888b	.789	.389	.191	.764**		
3	.932c	.869	.080	.067	.434**	.502**	
4	.956d	.914	.045	.025	.288*	.285*	.429**

P < .05. **p < .01; Seg = segmental features, Sseg = suprasegmental features, Phr = phrasing.

		ANOVA				
	Model	SS	df	MS	F	Sig.
1	Regression	261.327	1	261.327	19.328	.000a
	Residual Total	392.092	29	13.520		
		653.419	30			
2	Regression	515.584	2	257.792	52.368	.000b
	Residual Total	137.835	28	4.923		
		653.419	30			
3	Regression	568.088	3	189.363	59.917	.000c
	Residual Total	85.331	27	3.160		
		653.419	30			
4	Regression	597.391	4	149.348	69.306	.000d
	Residual Total	56.028	26	2.155		
		653.419	30			

(a) Predictors: (Constant), Grammar; (b) Predictors: (Constant), Grammar, Segmental features; (c) Predictors: (Constant), Grammar, Segmental features, Suprasegmental features; (d) Predictors: (Constant), Grammar, Segmental features, Suprasegmental features, Phrasing; (e) Dependent variable: L2 listening proficiency.

comprehension test. The proficient listeners' group consisted of 15 participants with test scores of 30 and above, and the less proficient listeners' group consisted of 16 participants whose scores were 29 points and below. **Table 5** presents mean scores and standard deviations for each variable by group.

Pearson product-moment correlation coefficients were conducted to examine how each of the five types of data was correlated with listening comprehension performance; **Table 6** and **Table 7** show the results. A remarkable contrast was found between the two groups; that is, while most of the explanatory variables, excepting only phrasing, revealed significant correlations with L2 listening ability among less proficient listeners, none indicated a significant correlation among proficient listeners.

5. Discussion

In order to answer the first research question, the intercorrelations between explanatory and dependent variables were first investigated among the participants as a whole. The results revealed that all the explanatory variables were significantly correlated to L2 listening ability. In particular, reading aloud performance, which was calcu-

Table 5. Descriptive statistics for proficient and less proficient listeners.

	Score range	Less proficient listeners (N = 16)		Proficient listeners (N = 15)	
		M	SD	M	SD
Listening	0 - 45	25.6	3.08	33.1	1.96
Reading aloud total	0 - 30	19.0	4.42	21.3	2.80
Segmental features	0 - 10	5.94	1.56	6.73	1.12
Suprasegmental features	0 - 10	6.25	1.75	7.07	1.00
Phrasing	0 - 10	6.81	1.37	7.53	1.14
Grammaticality judgment	0 - 45	17.6	6.50	23.5	5.34

Table 6. Bivariate correlations: Less proficient listeners (N = 16).

	Listening	Reading aloud			
		Total	Seg	Sseg	Phr
Reading aloud total	.93**				
Segmental features	.86**	.92**			
Suprasegmental features	.72**	.87**	.70**		
Phrasing	.93**	.94**	.80**	.77**	
Grammaticality judgment	.79**	.73**	.66**	.63**	.71**

Table 7. Bivariate correlations: Less proficient listeners (N = 16).

	Listening	Reading aloud			
		Total	Seg	Sseg	Phr
Reading aloud total	.95**				
Segmental features	.68**	.76**			
Suprasegmental features	.67**	.66**	.38		
Phrasing	.64*	.63*	.05	.28	
Grammaticality judgment	.01	-.06	-.07	.22	-.14

Note: **Correlation is significant at the .01 level (two-tailed). Seg = segmental features, Sseg = suprasegmental features, Phr = phrasing.

lated on the basis of the total scores for all three component measures, was found to be correlated to L2 listening, with an exceptionally high level of significance ($r = .95, p < .001$). The corresponding scatterplots in **Figure 2**, that is, (LIS, RA) and (RA, LIS), show that most of the markers are lined up in an orderly fashion on the regression line, indicating a nearly perfect linear fit between the two variables. Especially when one considers that there has been only a limited amount of previous literature examining the direct relationship between these variables, and that therefore there are only very few sources for comparison, this remarkably high correlation serves as a robust support for the idea of a close link between L2 speech production and perception.

A hierarchical regression analysis was then conducted to examine the degree to which L2 listening ability is explained by L2 grammatical knowledge and by each of the three reading aloud components. The results revealed that L2 grammatical knowledge explained 40% of the variance of L2 listening ability, and that after it was accounted for, each of the three components made significant contributions as well, jointly explaining 52% of the total variance. This strong connection between L2 speech production ability and L2 listening (e.g., [Walter, 2008](#); [Wilson & Iacoboni, 2006](#)) yields solid evidence that accurate and fluent L2 production ability contributes critically to L2 speech perception, particularly in light of the growing amount of empirical findings supporting a (partial) causal relationship between production and perception, for example, those showing impaired perception performance from disrupted motor activity (e.g., [Meister et al., 2007](#)), positive impact of non-word repetition practice on auditory activation (e.g., [Rauschecker, Pringle, & Watkins, 2008](#)), or performance improvement of phonological coding as a result of production practice (e.g., [Miyasako, 2008](#)), as well as from the wide range of studies suggesting that articulatory-to-auditory/-orosensory forward mapping (in terms of the level of cortical activations of the relevant brain regions) is a function of L2 proficiency (e.g., [Callan et al., 2004](#); [Gandour et al., 2007](#)).

However, there are a few novel points in the current results. One is related to the sheer predictive value of L2 grammatical knowledge, accounting for 40% of the total variance, in explaining L2 listening. This is much higher than the results obtained by previous studies, such as [Vandergrift \(2006\)](#), where L2 proficiency explained 25% of L2 listening comprehension performance, or [Mecartty \(2000\)](#), where it did not turn out to be a significant predictor for L2 listening. However, in prior L2 reading studies, findings on the shared variance between L2 knowledge and L2 reading have been comparable to those in the current study. For instance, the results of [Bernhardt and Kamil \(1995\)](#) showed that L2 proficiency accounted for 38% of variance, and [Lee and Schallert \(1997\)](#) also found it accounting for approximately 40%. Although a closer look into the relationship between L2 listening and reading is beyond the scope of the present study and research questions as they were conceived, this contrast could perhaps come from factors such as differences in the linguistic profiles of the participants: those in [Vandergrift \(2006\)](#) were learning French in an L2 setting in Canada, and some of them had been enrolled in French immersion programs or even had francophone parents, which may have helped reduce the impact of L2 grammatical knowledge compared to that seen in the current participants. Another candidate to explain the pronounced contribution of L2 grammatical knowledge in this study is the relationship between the participants' L2 proficiency in general and their score on the listening comprehension test in particular. As mentioned in the literature review section of this paper, the impact of general L2 proficiency on L2 reading ability is relative to factors such as the complexity of a reading text, and this pattern can certainly be applied to the present case: that is, the difficulty level of the listening materials used in this study (the practice TOEIC texts) may have been high enough to strain the participants' L2 proficiency. A third possible factor is related to the differences between specific L1-L2 relationships: whereas the participants in [Vandergrift \(2006\)](#) and [Mecartty \(2000\)](#) were L1 English speakers learning French or Spanish as an L2, those in the current study were L1 speakers of Japanese learning English, which leads us to speculate that greater L1-L2 distance may have contributed to the greater impact of L2 proficiency in this case (e.g., [Jarvis & Pavlenko, 2007](#); [Koda, 2008](#)). Indeed, the participants in the above-mentioned reading studies were either beginner level (in Bernhardt and Kamil's study) or had an L1-L2 distance much like the present participants (Lee and Schallert's study, which involved L1 Korean speakers learning English). These arguments are all relevant to the analysis of the second research question, and will be returned to later in this section.

After the predictive value of L2 grammatical knowledge, the second point of note in the findings, and one more crucially related to the aim of this study, is the finding that the predictive value of L2 grammatical knowledge became non-significant, though still marginal ($\text{sig.} = .083$), when the production performance of segmental features was entered into the regression. This can be explained by a range of empirical evidence from cognitive psychology suggesting that the efficiency of encoding and maintenance of phonological information bears a

causal relationship to overall language acquisition (e.g., Baddeley, Gathercole, & Papagno, 1998), on the assumption that accurate phonological representations certainly entail phonological awareness. In particular, a range of robust L2 research evidence has indicated a significant contribution of phonological awareness and phonological memory to the acquisition of vocabulary (e.g., Ellis & Sinclair, 1996; Service, 1992) and grammar (e.g., Robinson, 1997; Williams & Lavatt, 2005). These empirical findings support the assumption that the predictive power of pronunciation accuracy is already been reflected in the predictive value assigned to L2 grammatical knowledge in regression Model 1. The previous studies observing the impact of L2 proficiency on L2 listening performance could have provided a different picture if L2 pronunciation accuracy had been included as an explanatory variable; the current study serves as a precursor in this sense.

The third point of note is the magnitude of the predictive value of the production performance on individual speech sounds, which accounted for as much as 39% of the variance, underscoring the impact of accurate production of segmental features on L2 listening ability. Further speculation is required to explain this finding: for instance, are orthographic and phonological processing skills, which are presumably reflected in accurate and fluent production ability, involved here? Remember that production performance was measured using a reading aloud task, which seems to inevitably involve orthographic and phonological processing (knowledge of L2 grapheme-phoneme correspondences.) A body of empirical evidence has suggested that (visual-dependent) acquisition of L2 orthography does modify L2 phonological awareness (e.g., Morais, Bertelson, Cary, & Alegria, 1986) and therefore yields significant impact on on-line L2 speech processing (e.g., Dehaene et al., 2010; Nation & Hulme, 2011). This in turn indicates an effect of L2 orthographic experience on the development of L2 phonological awareness, and therefore, L2 speech perception, implying that efficient orthographic-phonological processing based on accurate letter-sound mapping is at least partially reflected in phoneme-level pronunciation accuracy, which may have amplified the predictiveness of production performance on segmental features for L2 listening.

Detailed consideration from a crosslinguistic perspective may provide further support for a crucial role played by the lower-level processing skills in the remarkable impact of the current component on L2 listening. The development of L2 speech production is closely linked to the acquisition of L2 phonological awareness (e.g., Price, 2012), and crosslinguistic studies focusing on the impact of transfer of the L1 sound system in the acquisition of the L2 speech perception have demonstrated that the development of L2 phonological awareness is crucially related to the overlap between pronunciation features in the listener's L1 and L2 (e.g., Jarvis & Pavlenko, 2007). Some studies report that participants had difficulty perceiving the distinction between sound pairs that were phonemic (contrastive) in the L2 but not the L1 (e.g., Aoyama, 2003; Escudero & Boersma, 2004). Japanese, the L1 of the current study's participants, has a mora-based rhythm with a consonant-vowel (CV) open syllable as the canonical syllable, while that of English is stress-timed and features a wide variety of both open syllables, including CV and V, and closed syllables, such as CVC, CCVC, and CCVCC (closed syllable) (see Laver, 1994; Ohata, 2004, for a detailed review). At the level of segmental features, while Japanese distinguishes only 5 vowels, (Standard US) English identifies roughly 15 of them including diphthongs (e.g., Mannell, Cox, & Harrington, 2009). Other dialects may recognize more: Roach, 2004, for instance, identifies 11 monophthongs and 19 diphthongs in Received Pronunciation); and with regard to consonants, English has a considerably wider variety of fricatives and affricatives than Japanese. It is reasonable to assume that the distinctions between these sounds are difficult for Japanese L1 learners to perceive. Although several other factors likely affect crosslinguistic transfer (e.g., phonetic environment, universal phonological constraints), L1 - L2 distance in terms of (supra-) segmental sound system may yield the most significant impact on the efficiency of learning and development of L2 phonological representations (e.g., Hancin-Bhatt & Bhatt, 1997), which may be a crucial explanatory factor for the pronounced predictive value of the production performance of the individual speech sounds found in the current study.

Returning to the regression analysis, each of the other two components—the production performance of suprasegmental features and phrasing—was also found to be a significant predictor of L2 listening ability, though less so than that of segmental features, with the former explaining an additional 8% of shared variance, and the latter a further 5%. This finding indicates that the suprasegmental production efficiency, including intonation, sentence-level stress, and rhythm, and the accurate segmenting/phrasing in production significantly contributed to L2 listening. However, also noteworthy is that when these components were added to the model, the level of predictiveness of L2 grammatical knowledge for L2 listening became non-significant (i.e., sig. = .465, for fluency [Model 3] and sig. = .740, for segmentation [Model 4]). Therefore, it seems reasonable to assume that, since

both of these two components (the production performance of suprasegmental features and phrasing) tap both semantic and syntactic processing (e.g., Nuttall, 2005), a large part of the explanatory value of both components, as well as the segmental feature component, had been incorporated in the L2 grammatical knowledge variable. That is, the predictive power that initially seemed to be explained by L2 knowledge proved in fact to be accounted for, step by step, by the three components as each was entered in the regression.

Regarding the second research question, that is, how the respective contributions from the explanatory variables will change depending on L2 listening proficiency, the result revealed a distinctive contrast between the two participants (which were defined by their L2 listening comprehension performance). Specifically, while highly significant intercorrelations were maintained for all variables among less proficient listeners (see Table 6), more than half these correlations (8 out of 15) lost significance among the proficient listeners (see Table 7). Even when significance was maintained, it was at lower levels, except for listening and reading aloud. Perhaps the most important point here is that while L2 grammatical knowledge was highly significantly correlated with all other variables in the less proficient group, all these intercorrelations completely disappeared in the proficient group. (This contrast can be observed in the rightmost and bottom scatterplots in Figure 2, all of which visualize similar relationships between grammatical knowledge and the other variables: for each variable, the loess curves become almost flat among proficient performers). The finding that the correlations between listening and reading aloud as well as the segmental and suprasegmental feature components remained highly significant among proficient listeners indicates that these variables remain closely linked even as listening proficiency improves, suggesting that pronunciation accuracy and fluency still play an important role in explaining L2 listening ability among fairly advanced listeners. The non-significance of the correlations between L2 grammatical knowledge and listening and reading aloud performance among proficient listeners may possibly be explained by the linguistic threshold for L2 listening/reading ability (e.g., Vandergrift, 2006, for listening; e.g., Bossers, 1991, for reading). That is, once L2 listeners/readers have attained a certain level of L2 knowledge, the predictive powers of these correlations may no longer be significant. The finding of Mearcy (2000) wherein L2 grammatical knowledge was not a significant predictor of L2 listening ability is also compatible with this argument. Mearcy indeed argues that the reason for the non-significant contribution of L2 grammar knowledge to L2 listening in her study could have been that her participants were college students, with enough listening proficiency to mask the contribution of L2 knowledge (p. 337).

Thus, the overall predictiveness of L2 knowledge for L2 listening (explaining as much as 40% of the total variance) originated from the highly correlated relationship between these variables observed among less proficient listeners. Therefore, the argument that factors such as L1-L2 distance or the interaction between L2 proficiency and the difficulty of listening material may have amplified the impact of L2 grammatical knowledge can only apply to the less proficient group; in contrast, the proficient participants may have reached a level where the impact of L2 knowledge is negligible. However, it is important to note that the three components of reading aloud are still significantly correlated with listening ability, suggesting that pronunciation accuracy and fluency still matter for this group. Additional support for this argument comes from the finding that the significance of the phrasing component becomes lower (.64*, $p < .05$) than the other two components (.68**, $p < .01$, for segmental features and .67**, $p < .01$, for suprasegmental features) among proficient listeners, indicating that phrasing may be more closely linked to grammatical knowledge than the other two components.

The present study attempted to examine the relationship between reading aloud (RA) performance, grammatical knowledge, and listening ability in a sample of 31 college-level Japanese EFL learners. The results demonstrated highly significant intercorrelations between all the explanatory variables—grammatical knowledge and the three components of reading aloud—and listening ability. The subsequent regression analyses showed remarkably high contributions to listening from all the variables, with production accuracy of segmental features showing an especially pronounced impact. Supplementary analyses dividing the participants by listening ability demonstrated a remarkable contrast: while on the one hand, a significant correlation was maintained between grammatical knowledge and listening among less proficient listeners but disappeared among more proficient listeners; however, on the other hand, a significant correlation between RA and listening performance was maintained in both groups, indicating that production accuracy/fluency still plays an important role in L2 listening at an advanced level.

6. Conclusion

Newborns start to smile at their caregivers around one month after birth (e.g., Bertin & Striano, 2006). Initially,

their smiles are just an innate motoric imitation of maternal smiles; only later do they become associated with a feeling of pleasure through face-to-face interaction over mutual contingent smiles with their caregivers (though exactly when that transition occurs has yet to be clarified; e.g., Wörmann, Holodynski, Kärtner, & Keller, 2012). This “motor imitation first” paradigm can also be observed in verbal behaviors of newborns, as evidenced by (inter alia) Mampe, Friederici, Christophe, & Wermke (2009), who found that newborns’ cry melody was already influenced by native language intonation patterns in the first week of life, and Chen, Striano, & Rakoczy (2004), who found that newborns of 1-7 days old were able to perform mouth movements corresponding to both vowel and consonant vocal models, both of which suggest closely linked innate auditory-articulatory mapping.

In both L1 and L2 research, explanations of causality in language acquisition have traditionally been based on the mostly tacit consensus that “listening is the basis for overall [...] proficiency” (e.g., Krashen & Terrell, 1983; Feyten, 1991). L2 studies investigating the impact of perceptual training on production suggest that improvement in L2 listening perception can indeed serve as a “pre-requisite or co-requisite to making connections to meaning and use” (Derwing et al., 2012: p. 252) and thus to more accurate productions (e.g., Derwing et al., 2012; Thomson, 2012). However, few of these studies have referred to the possible involvement of subvocal repetition or rehearsal during or after perception instruction, which could by no means have been ruled out as a factor unless, for instance, the training had been conducted under a condition of irrelevant background speech noises, which has been evidenced to impair verbal recall performance by interfering with the maintenance of auditorily presented target stimuli (e.g., Salamé & Baddeley, 1982; Kato, 2009a); or under a condition of articulatory suppression, which requires repeated articulation of irrelevant speech sounds (such as the words *blah* or *the*) during a serial memory task, preventing the participants from rehearsing items to be recalled in memory (e.g., Levy, 1971; Kato, 2009b). The present study thus attempted to investigate whether the paradigm of “production facilitates perception” holds water in the context of the acquisition of L2 listening ability by investigating how L2 reading aloud performance, along with L2 grammatical knowledge, is related to L2 listening ability.

The first correlational analyses of the participants as a whole and the subsequent multiple regression analyses aimed to examine the first research question. The results of the correlational analyses demonstrated highly significant intercorrelations between all the explanatory variables—L2 grammatical knowledge and three components of reading aloud as well as the total reading aloud score—and L2 listening ability. The regression analyses in fact showed remarkably high contributions from all the explanatory variables to L2 listening, with production performance of individual speech sounds showing an especially pronounced impact. Based on previous findings on the relationship between phonological awareness and overall L2 acquisition (e.g., Service, 1992; Williams & Lavatt, 2005), it was also suggested that pronunciation accuracy of segmental features, along with that of suprasegmental features and phrasing (the other two components), partly explained the role of L2 grammatical knowledge in total variance of L2 listening.

The second correlational analysis, with the participants divided into two groups on the basis of L2 listening test scores, demonstrated a remarkable contrast in terms of the level of correlation between L2 grammatical knowledge and L2 listening ability. That is, while high significance was maintained between these variables among less proficient listeners, the relationship was non-significant among the (more) proficient listeners, suggesting that the latter group’s L2 knowledge had reached a level where it no longer affected L2 listening performance, whereas that of less proficient listeners still made a significant contribution to L2 listening ability. In contrast, significant correlations were maintained among proficient listeners between the three components of reading aloud and listening ability, as was the highly significant relationship between reading aloud total score and listening ability, indicated that pronunciation accuracy and fluency played an important role in L2 listening ability for both groups.

These findings provide robust support for the argument that the establishment of pronunciation accuracy/fluency is crucial for the development of listening ability and that this impact of production ability may linger to a fairly advanced stage of L2 listening learning, in particular as a function of factors such as participants’ L1 - L2 relationship and the relationship between their L2 proficiency and the familiarity and difficulty of listening materials. Moreover, the differences in significance between the three components of reading aloud mean that the acquisition of pronunciation accuracy seems paramount to successful development of L2 listening ability. This supposition fits well with the concept of the “bottom-up primacy” of successful listening conditions proposed in an L1 context by Marslen-Wilson (1989), where accurate individual sound perception played a critical role in narrowing down potential candidate words in the mental lexicon while listening. Wilson (2003) also argued for an important role of bottom-up processing in L2 listening, mentioning that although utilizing compen-

satory top-down strategies is widely encouraged across all proficiency levels of L2 learners, the ultimate goal of the development of L2 listening ability is to rely less on contextual guesswork and more on hearing what was actually said, further indicating that “better bottom-up processing ought to lead in turn to better top-down processing, and that teaching should reflect this” (p. 342). The proposal of “bottom-up primacy” by Marslen-Wilson (1989) is compatible with the theoretical framework presented by Robinson (2005), which assumed a hierarchical relationship, using concentric circles, between changes in the relative contribution of various aptitude factors to the development of overall L2 acquisition. Robinson placed ability factors such as processing speed, phonological working memory capacity, phonological working memory speed, and rote memory in the inner circle of the diagram, meaning that they make an initial contribution to L2 acquisition, after which groups of peripheral factors such as task aptitude and pragmatic/interactional abilities/traits may affect overall development (p. 52). Hulstijn (2011) proposed a definition of *language proficiency* (LP) in which the phonetic-phonological, morphophonological, morphosyntactic, and lexical domains form the *core components* of LP, whereas areas of a less linguistic or non-linguistic nature, such as strategic or metacognitive abilities, constitute *peripheral components* (p. 242). These arguments are in line with the idea of an “acquisition hierarchy” in L2 listening development that is adopted in the present study, whereby the acquisition of pronunciation accuracy should come first as a basis for the development of higher-level processing ability.

The findings of the current study have important implications regarding the usefulness of reading aloud activities in the development of L2 listening. Kato (2012) provided a range of reading aloud activities, such as *choral/unison reading*, in which learners read a text aloud, together, at the same speed, following a model reading; *individual reading*, in which learners read the text aloud at their own speed; and other variations such as *overlapping* or *synchronized reading*, which requires that learners read the text aloud as they listen to the model reading and try to keep up as accurately as possible. The *shadow talk* technique, in which learners “speak out” a text without access to a script, again following a model reading, has also been reported to be useful in establishing an L2 sound repertoire (e.g., Mochizuki, 2004).

Future research should adopt a longitudinal design in order to focus on the cause-effect relationship between long-term reading aloud practice and development of L2 listening ability, possibly with a larger amount of participants than the present study and certainly with more varied L2 proficiency and L2 learning contexts. Research to examine the long-term effect of practice can also incorporate more refined test designs so as to look at the relationship between programs for training in particular (supra-) segmental articulations and the improvement of perception ability.

Given the high levels of intercorrelation found in the present study between reading aloud performance, grammatical knowledge, and listening ability, as well as the previously evidenced strong link between reading aloud performance and reading ability (e.g., Miyasako, 2008), future research should also explore the potential of the reading aloud technique as an effective and practical assessment tool to measure overall L2 proficiency, especially for the purpose of screening and placement in various L2 research/educational settings. Various types of assessment tools are used for the measurement of L2 proficiency in the research arena of SLA, from the major testing systems (e.g., for English, IELTS, TOEFL, and TOEIC), testing tools in particular linguistic fields (such as the Peabody Picture Vocabulary Test and Nation’s Vocabulary Level Test), and more general testing techniques such as the cloze test and C-test (see Hulstijn, 2012, for a comprehensive review of the measurement of language proficiency). Elicited Imitation Test (EIT) and Elicited Oral Response (EOR) testing have also been proposed as effective assessment tools, particularly for L2 oral proficiency (e.g., Vinther, 2002; Cox & Davies, 2012; Wu & Ortega, 2013). In contrast to EIT and EOR, which examine correctness of repetitions of the target sentences (on a single four-point correctness scale in the former test, and by the number of automatically recognized words in the latter), the scoring procedure for reading aloud performance in the current study is more fine-grained, tapping three different components (the segmental feature component, to focus on individual sounds; the suprasegmental feature component, for rhythm, sentential stress, and intonation; and the phrasing component), which have been shown to reflect L2 grammatical knowledge as well as the accuracy and fluency of segmental and suprasegmental features. Furthermore, the reading aloud processing obligatorily involves basic component skills such as phonological and orthographic processing skills, which have been empirically shown to play important roles in reading development (e.g., Kato, 2009b; see also Nassaji, 2014, for what is currently the latest comprehensive review). These features indicate the great potential of the reading aloud technique as an effective, viable, and practical assessment tool to measure overall L2 proficiency covering listening, reading, and speaking, as well as underlying L2 grammatical knowledge. Future extensions of this research with further

refined designs and more varied population profiles in more varied educational settings are needed to further articulate the tool's possible utility in a way that is valid and reliable.

To sum up, the present study attempted to examine a twofold research question: 1) *How are L2 reading aloud performance, as measured in terms of three components—the production performance of segmental features, suprasegmental features, and phrasing—and L2 grammatical knowledge, as measured by the grammaticality judgment task, related to L2 listening performance?* and 2) *Do the relationships found in the first research question, if any, reveal any change as a function of L2 listening proficiency?* Regarding the first research question, the results of the correlational analyses demonstrated highly significant intercorrelations between all the explanatory variables—grammatical knowledge and three components of reading aloud—and listening ability. The subsequent regression analyses showed remarkably high contributions from all the explanatory variables to listening ability, with production accuracy of segmental features showing an especially pronounced impact. The second correlational analysis, with the participants divided into two groups on the basis of listening test scores, was carried out to examine the second research question, and the result demonstrated a remarkable contrast. That is, while a significant correlation was maintained between grammatical knowledge and listening among less proficient listeners, it disappeared among more proficient listeners. In contrast, a significant correlation between RA and listening performance was maintained in both groups, indicating that production accuracy/fluency still plays an important role in L2 listening at an advanced level. These findings provide robust support for the argument that the establishment of pronunciation accuracy/fluency is crucial for the development of listening ability, even at a fairly advanced stage of L2 listening learning.

References

- Al-jasser, F. (2008). The Effect of Teaching English Phonotactics on the Lexical Segmentation of English as a Foreign Language. *System*, 36, 94-106. <http://dx.doi.org/10.1016/j.system.2007.12.002>
- Aoyama, K. (2003). Perception of Syllable-Initial and Syllable-Final Nasals in English by Korean and Japanese Speakers. *Second Language Research*, 19, 251-265. <http://dx.doi.org/10.1191/0267658303sr222oa>
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511667350>
- Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.
- Baddeley, A. D. (2000). The Episodic Buffer: A New Component of Working Memory? *Trends in Cognitive Sciences*, 4, 417-423. [http://dx.doi.org/10.1016/S1364-6613\(00\)01538-2](http://dx.doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, 105, 158-173. <http://dx.doi.org/10.1037/0033-295X.105.1.158>
- Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting Relationships between L1 and L2 Reading: Consolidating the Linguistic Threshold and the Linguistic Interdependence Hypotheses. *Applied Linguistics*, 16, 15-34. <http://dx.doi.org/10.1093/applin/16.1.15>
- Bertin, E., & Striano, T. (2006). The Still-Face Response in Newborn, 1.5-, and 3-Month-Old Infants. *Infant Behavior and Development*, 29, 294-297. <http://dx.doi.org/10.1016/j.infbeh.2005.12.003>
- Birch, B. M. (2007). *English L2 Reading: Getting to the Bottom* (2nd ed.). Nahwah, NJ: Lawrence Erlbaum.
- Bossers, B. (1991). On Thresholds, Ceilings and Short-Circuits: The Relation between L1 Reading, L2 Reading and L2 Knowledge. *AILA Review*, 8, 45-60.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic Perceptual Identification by Native- and Second-Language Speakers Differentially Activates Brain Regions Involved with Acoustic Phonetic Processing and Those Involved with Articulatory-Auditory/Orosensory Internal Models. *NeuroImage*, 22, 1182-1194. <http://dx.doi.org/10.1016/j.neuroimage.2004.03.006>
- Carrell, P. L. (1991). Second Language Reading: Reading Ability or Language Proficiency? *Applied Linguistics*, 12, 159-179. <http://dx.doi.org/10.1093/applin/12.2.159>
- Casserly, E. D., & Pisoni, D. B. (2010). Speech Perception and Production. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 629-647. <http://dx.doi.org/10.1002/wcs.63>
- Chee, M. W., Hon, N., Lee, H. L., & Soon, C. S. (2001). Relative Language Proficiency Modulates BOLD Signal Change When Bilinguals Perform Semantic Judgments. *NeuroImage*, 13, 1155-1163. <http://dx.doi.org/10.1006/nimg.2001.0781>
- Chen, X., Striano, T., & Rakoczy, R. (2004). Auditory-Oral Matching Behavior in Newborns. *Developmental Science*, 7, 42-47. <http://dx.doi.org/10.1111/j.1467-7687.2004.00321.x>

- Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., & Pesaran, B. (2014). Sensory-Motor Transformations for Speech Occur Bilaterally. *Nature*, *507*, 94-98. <http://dx.doi.org/10.1038/nature12935>
- Cox, T., & Davies, R. S. (2012). Using Automatic Speech Recognition Technology with Elicited Oral Response Testing. *CALICO Journal*, *29*, 601-618. <http://dx.doi.org/10.11139/cj.29.4.601-618>
- Crosson, A. C., & Lesaux, N. K. (2010). Revisiting Assumptions about the Relationship of Fluent Reading to Comprehension: Spanish-Speakers' Text-Reading Fluency in English. *Reading and Writing*, *23*, 475-494. <http://dx.doi.org/10.1007/s11145-009-9168-8>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Cohen, L. et al. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, *330*, 1359-1364. <http://dx.doi.org/10.1126/science.1194140>
- Derwing, T. M., Thomson, R. I., Foote, J. A., & Munro, M. J. (2012). A Longitudinal Study of Listening Perception in Adult Learners of English: Implications for Teachers. *Canadian Modern Language Review*, *68*, 247-266. <http://dx.doi.org/10.3138/cmlr.1215>
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding Motor Events: A Neurophysiological Study. *Experimental Brain Research*, *91*, 176-180. <http://dx.doi.org/10.1007/BF00230027>
- Ellis, N. C., & Sinclair, S. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *Quarterly Journal of Experimental Psychology Section A*, *49*, 234-250. <http://dx.doi.org/10.1080/713755604>
- Escudero, P., & Boersma, P. (2004). Bridging the Gap between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, *26*, 551-585. <http://dx.doi.org/10.1017/S0272263104040021>
- Farnia, F., & Geva, E. (2011). Cognitive Correlates of Vocabulary Growth in English Language Learners. *Applied Psycholinguistics*, *32*, 711-738. <http://dx.doi.org/10.1017/S0142716411000038>
- Feyten, C. (1991). The Power of Listening Ability: An Overlooked Dimension in Language Acquisition. *Modern Language Journal*, *75*, 173-180. <http://dx.doi.org/10.1111/j.1540-4781.1991.tb05348.x>
- Field, A. (2005). *Discovering Statistics Using SPSS* (2nd ed.). London: Sage.
- Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511575945>
- French, L. (2003). *Phonological Working Memory and L2 Acquisition: A Developmental Study of Quebec Francophone Children Learning English*. Ph.D. Thesis, Québec: Université Laval.
- Gandour, J., Tong, Y., Talavage, T., Wong, D., Dziedzic, M., Xu, Y., Li, X., & Lowe, M. (2007). Neural Basis of First and Second Language Processing of Sentence-Level Linguistic Prosody. *Human Brain Mapping*, *28*, 94-108. <http://dx.doi.org/10.1002/hbm.20255>
- Gibson, S. (2008). Reading Aloud: A Useful Learning Tool? *ELT Journal*, *62*, 29-36. <http://dx.doi.org/10.1093/elt/ccm075>
- Goh, C. (2000). A Cognitive Perspective on Language Learners' Listening Comprehension Problems. *System*, *28*, 55-75. [http://dx.doi.org/10.1016/S0346-251X\(99\)00060-3](http://dx.doi.org/10.1016/S0346-251X(99)00060-3)
- Graham, S. (2006). Listening Comprehension: The Learners' Perspective. *System*, *34*, 165-182. <http://dx.doi.org/10.1016/j.system.2005.11.001>
- Hancin-Bhatt, B., & Bhatt, R. (1997). Optimal L2 Syllables: Interactions of Transfer and Developmental Effects. *Studies in Second Language Acquisition*, *19*, 331-378. <http://dx.doi.org/10.1017/S0272263197003033>
- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI Study of Bilingual Sentence Comprehension and Workload. *NeuroImage*, *15*, 647-660. <http://dx.doi.org/10.1006/nimg.2001.1001>
- Heiberger, R. M., & Holland, B. (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R and SAS*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4757-4284-8>
- Hulstijn, J. H. (2001). Intentional and Incidental Second Language Vocabulary Learning: A Reappraisal of Elaboration, Rehearsal and Automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258-286). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524780.011>
- Hulstijn, J. H. (2003). Connectionist Models of Language Processing and the Training of Listening Skills with the Aid of Multimedia Software. *Computer Assisted Language Learning*, *16*, 413-425. <http://dx.doi.org/10.1076/call.16.5.413.29488>
- Hulstijn, J. H. (2011). Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment. *Language Assessment Quarterly*, *8*, 229-249. <http://dx.doi.org/10.1080/15434303.2011.565844>
- Iacoboni, M. (2008). The Role of Premotor Cortex in Speech Perception: Evidence from fMRI and rTMS. *Journal of Physiology-Paris*, *102*, 31-34. <http://dx.doi.org/10.1016/j.jphysparis.2008.03.003>
- Ichitani, Y. (Ed.) (2011). *Unicorn English Course I: New Edition*. Tokyo: Bun-Eido.
- Iwamura, K., & Smillie, B. (2007). *New Edition of TOEIC Test Trial Examination with 600 Questions*. Tokyo: ALC.

- Janzen, J. (2007). Preparing Teachers of Second Language Reading. *TESOL Quarterly*, 41, 707-729.
- Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.
- Kato, S. (2009a). Interaction between Processing and Maintenance in Online L2 Sentence Comprehension: Implications for Linguistic Threshold Hypothesis. *Asian EFL Journal*, 11, 235-273.
- Kato, S. (2009b). Suppressing Inner Speech in ESL Reading: Implications for Developmental Changes in Second Language Word Recognition Processes. *Modern Language Journal*, 93, 471-488.
<http://dx.doi.org/10.1111/j.1540-4781.2009.00926.x>
- Kato, S. (2012). Bridging Theory and Practice: Developing Lower-Level Processing Skills in L2 Reading. *The Language Learning Journal*, 40, 193-206. <http://dx.doi.org/10.1080/09571736.2011.621168>
- Kiany, G. R., & Shiramiry, E. (2002). The Effect of Frequent Dictation on the Listening Comprehension Ability of Elementary EFL Learners. *TESL Canada Journal*, 20, 57-63.
- Kincaid, J. P., Fishburn, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of New Readability Formulas for Navy Enlisted Personnel* (pp. 8-75). Technical Report Research Branch Report, Millington, TN: Naval Technical Training. Memphis, TN: US Naval Air Station.
- Koda, K. (2008). Impacts of Prior Literacy Experience on Second-Language Learning to Read. In K. Koda, & A. M. Zehler (Eds.), *Learning to Read across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development* (pp. 68-96). Mahwah, NJ: Erlbaum.
- Koda, K., & Zehler, A. (Eds.). (2008). *Learning to Read across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development*. New York: Routledge.
- Krashen, S., & Terrell, T. (1983). *The Natural Approach: Language Acquisition in the Classroom*. Oxford: Pergamon Press.
- LaFrance, A., & Gottardo, A. (2005). A Longitudinal Study of Phonological Processing Skills and Reading in Bilingual Children. *Applied Psycholinguistics*, 26, 559-578. <http://dx.doi.org/10.1017/S0142716405050307>
- Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York: Routledge.
- Larson-Hall, J., & Herrington, R. (2010). Improving Data Analysis in Second Language Acquisition by Utilizing Modern Developments in Applied Statistics. *Applied Linguistics*, 31, 368-390. <http://dx.doi.org/10.1093/applin/amp038>
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9781139166621>
- Lee, J., & Schallert, D. L. (1997). The Relative Contribution of L2 Language Proficiency and L1 Reading Ability to L2 Reading Performance: A Test of the Threshold Hypothesis in an EFL Context. *TESOL Quarterly*, 31, 713-739.
<http://dx.doi.org/10.2307/3587757>
- Levy, B. A. (1971). Role of Articulation in Auditory and Visual Short-Term Memory. *Journal of Verbal Learning and Verbal Behavior*, 10, 123-132. [http://dx.doi.org/10.1016/S0022-5371\(71\)80003-8](http://dx.doi.org/10.1016/S0022-5371(71)80003-8)
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the Speech Code. *Psychological Review*, 74, 431-461. <http://dx.doi.org/10.1037/h0020279>
- Lieberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception Revised. *Cognition*, 21, 1-36.
[http://dx.doi.org/10.1016/0010-0277\(85\)90021-6](http://dx.doi.org/10.1016/0010-0277(85)90021-6)
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Del Gratta, C., Romani, G., & Belardinelli, M. O. (2010). Sensory-Motor Brain Network Connectivity for Speech Comprehension. *Human Brain Mapping*, 31, 567-580.
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, 19, 1994-1997. <http://dx.doi.org/10.1016/j.cub.2009.09.064>
- Mannell, R., Cox, F., & Harrington, J. (2009). *An Introduction to Phonetics and Phonology*. Macquarie University.
<http://clas.mq.edu.au/speech/phonetics/index.html>
- Marslen-Wilson, W. D. (1989). Access and Integration: Projecting Sound onto Meaning. In W. D. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 3-24). Cambridge, MA: MIT Press.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential Role of Premotor Cortex in Speech Perception. *Current Biology*, 17, 1692-1696. <http://dx.doi.org/10.1016/j.cub.2007.08.064>
- Mecartty, F. (2000). Lexical and Grammatical Knowledge in Reading and Listening Comprehension by Foreign Language Learners of Spanish. *Applied Language Learning*, 11, 323-348.
- Miyasako, N. (2008). Is the Oral Reading Hypothesis Valid? *Language Education and Technology*, 45, 15-34.
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy Training and Speech Segmentation. *Cognition*, 24, 45-64.
[http://dx.doi.org/10.1016/0010-0277\(86\)90004-1](http://dx.doi.org/10.1016/0010-0277(86)90004-1)
- Mochizuki, H. (2004). *Application of Shadowing to TEFL in Japan: The Case of Junior High School Students*. MA Thesis, Hyogo: Kwansai Gakuin University.

- Nassaji, H. (2014). The Role and Importance of Lower-Level Processes in Second Language Reading. *Language Teaching*, 47, 1-37. <http://dx.doi.org/10.1017/S0261444813000396>
- Nation, I. S. P. (2009). *Teaching ESL/EFL Reading and Writing*. New York: Routledge.
- Nation, K., & Hulme, C. (2011). Learning to Read Changes Children's Phonological Skills: Evidence from a Latent Variable Longitudinal Study of Reading and Nonword Repetition. *Developmental Science*, 14, 649-659. <http://dx.doi.org/10.1111/j.1467-7687.2010.01008.x>
- Nuttall, C. (2005). *Teaching Reading Skills in a Foreign Language*. Oxford: Macmillan Education.
- Oberman, L. M., Winkelman, P., & Ramachandran, V. S. (2007). Face to Face: Blocking Facial Mimicry Can Selectively Impair Recognition of Emotional Expressions. *Social Neuroscience*, 2, 167-178. <http://dx.doi.org/10.1080/17470910701391943>
- Ohata, K. (2004). Phonological Differences between Japanese and English: Several Potentially Problematic Areas of Pronunciation for Japanese ESL/EFL Learners. *Asian EFL Journal*, 6, 1-19.
- Osada, N. (2001). What Strategy Do Less Proficient Learners Employ in Listening Comprehension? A Reappraisal of Bottom-Up and Top-Down Processing. *Journal of Pan-Pacific Association of Applied Linguistics*, 5, 73-90.
- Price, C. J. (2012). A Review and Synthesis of the First 20 Years of PET and fMRI Studies of Heard Speech, Spoken Language and Reading. *NeuroImage*, 62, 816-847. <http://dx.doi.org/10.1016/j.neuroimage.2012.04.062>
- Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in Neural Activity Associated with Learning to Articulate Novel Auditory Pseudowords by Covert Repetition. *Human Brain Mapping*, 29, 1231-1242. <http://dx.doi.org/10.1002/hbm.20460>
- Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience*, 27, 169-192. <http://dx.doi.org/10.1146/annurev.neuro.27.070203.144230>
- Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain: How Our Minds Share Actions, Emotions, and Experience*. New York: Oxford University Press.
- Roach, P. (2004). British English: Received Pronunciation. *Journal of the International Phonetic Association*, 34, 239-245. <http://dx.doi.org/10.1017/S0025100304001768>
- Robinson, P. (1997). Individual Differences and the Fundamental Similarity of Implicit and Explicit Adult Second Language Learning. *Language Learning*, 47, 45-99. <http://dx.doi.org/10.1111/0023-8333.21997002>
- Robinson, P. (2005). Aptitude and Second Language Acquisition. *Annual Review of Applied Linguistics*, 25, 46-73. <http://dx.doi.org/10.1017/S0267190505000036>
- Rost, M. (2002). *Teaching and Researching Listening*. Harlow: Longman.
- Salamé, P., & Baddeley, A. D. (1982). Disruption of Short-Term Memory by Unattended Speech: Implications for the Structure of Working Memory. *Journal of Verbal Learning & Verbal Behavior*, 21, 150-164. [http://dx.doi.org/10.1016/S0022-5371\(82\)90521-7](http://dx.doi.org/10.1016/S0022-5371(82)90521-7)
- Segalowitz, N. (2003). Automaticity and Second Languages. In C. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382-408). Malden, MA: Blackwell.
- Service, E. (1992). Phonology, Working Memory, and Foreign-Language Learning. *Quarterly Journal of Experimental Psychology Section A*, 45, 21-50. <http://dx.doi.org/10.1080/14640749208401314>
- Shichino, M. (2006). Effects of Repeated Oral Reading Practice on the Reproduction of Key Words and Phrases. *46th Annual Conference of the Japan Association for Language Education and Technology*, Kyoto, 3 August 2006, 103-109.
- Shimizu, M. (2009). An Examination of Reliability of an Oral Reading Test for EFL Learners Using Generalizability Theory. *ARELE: Annual Review of English Language Education in Japan*, 20, 181-190.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to Talking Faces: Motor Cortical Activation during Speech Perception. *NeuroImage*, 25, 76-89. <http://dx.doi.org/10.1016/j.neuroimage.2004.11.006>
- Suzuki, J. (1998). *A Reprisal of Reading Aloud: An Empirical Study on the Effectiveness of Reading Aloud* (pp. 13-28). LET Kansai Chapter Collected Papers 7, Kobe, Hyogo: Rokko-Shuppan.
- Taillefer, G. F. (1996). L2 Reading Ability: Further Insight into the Short-Circuit Hypothesis. *The Modern Language Journal*, 80, 461-477. <http://dx.doi.org/10.1111/j.1540-4781.1996.tb05466.x>
- Thomson, R. I. (2012). Improving L2 Learners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, 62, 1231-1258. <http://dx.doi.org/10.1111/j.1467-9922.2012.00724.x>
- Vandergrift, L. (2003). From Prediction through Reflection: Guiding Students: Through the Process of L2 Listening. *The Canadian Modern Language Review*, 59, 425-440. <http://dx.doi.org/10.3138/cmlr.59.3.425>
- Vandergrift, L. (2006). Second Language Listening: Listening Ability or Language Proficiency? *The Modern Language Journal*, 90, 6-18. <http://dx.doi.org/10.1111/j.1540-4781.2006.00381.x>
- Vandergrift, L. (2011). Second Language Listening: Presage, Process, Product and Pedagogy. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (2nd ed., pp. 455-471). New York: Routledge.

- Vinther, T. (2002). Elicited Imitation: A Brief Overview. *International Journal of Applied Linguistics*, 12, 54-73. <http://dx.doi.org/10.1111/1473-4192.00024>
- Walter, C. (2008). Phonology in Second Language Reading: Not an Optional Extra. *TESOL Quarterly*, 42, 455-474.
- Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI Evidence for Cortical Modification during Learning of Mandarin Lexical Tone. *Journal of Cognitive Neuroscience*, 15, 1019-1027. <http://dx.doi.org/10.1162/089892903770007407>
- Williams, J. N., & Lavatt, P. (2005). Phonological Memory and Rule Learning. *Language Learning*, 55, 177-233. <http://dx.doi.org/10.1111/j.0023-8333.2005.00298.x>
- Wilson, M. (2003). Discovery Listening: Improving Perceptual Processing. *ELT Journal*, 57, 335-343. <http://dx.doi.org/10.1093/elt/57.4.335>
- Wilson, S. M., & Iacoboni, M. (2006). Neural Responses to Non-Native Phonemes Varying in Producibility: Evidence for the Sensorimotor Nature of Speech Perception. *NeuroImage*, 33, 316-325. <http://dx.doi.org/10.1016/j.neuroimage.2006.05.032>
- Wörmann, V., Holodyski, M., Kärtner, J., & Keller, H. (2012). A Cross-Cultural Comparison of the Development of the Social Smile: A Longitudinal Study of Maternal and Infant Imitation in 6- and 12-Week-Old Infants. *Infant Behavior and Development*, 35, 335-347. <http://dx.doi.org/10.1016/j.infbeh.2012.03.002>
- Wu, S. L., & Ortega, L. (2013). Measuring Global Oral Proficiency in SLA Research: A New Elicited Imitation Test of L2 Chinese. *Foreign Language Annals*, 46, 680-704. <http://dx.doi.org/10.1111/flan.12063>
- Xue, G., Dong, Q., Jin, Z., & Chen, C. (2004). Mapping of Verbal Working Memory in Nonfluent Chinese-English Bilinguals with Functional MRI. *NeuroImage*, 22, 1-10. <http://dx.doi.org/10.1016/j.neuroimage.2004.01.013>

Appendix: Text for Reading Aloud Task

Natsume Soseki went to Britain about 100 years ago as a student sent by the Ministry of Education. He was 33 years old and a professor at the Fifth High School in Kumamoto then. He left Yokohama by ship in September 1900, and reached London two months later. Britain was more developed than other countries in those days. There was already a web of undergrounds in London—30 years before the first underground in Tokyo. Everything Soseki saw and heard was a wonder. He enjoyed buying used books, walking in the parks and going to the theaters. He wrote this to his wife: “I wish you could see the wonderful theater shows. In one show, I saw about sixty women dancing on the stage in gorgeous costumes.”