All Theses and Dissertations

2018-03-28

Perceptual Proficiency Ratings of Obstruent Productions in L2 Learners of English as a Function of Speech Task Type, Word Position, and Listener Expertise

Rachel McPherson Zitting

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Communication Commons, and the Communication Sciences and Disorders Commons

Perceptual Proficiency Ratings of Obstruent Productions in L2 Learners of English

as a Function of Speech Task Type, Word Position, and Listener Expertise

Rachel McPherson Zitting

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Shawn L. Nissen, Chair
K. James Hartshorn
Christopher Dromey

Department of Communication Disorders

Brigham Young University

ABSTRACT

Perceptual Proficiency Ratings of Obstruent Productions in L2 Learners of English
as a Function of Speech Task Type, Word Position, and Listener Expertise

Rachel McPherson Zitting
Department of Communication Disorders, BYU
Master of Science

Second language (L2) learners of English must learn to produce English phonemes, words, and sentences. These L2 learners make many errors when learning English; they may change the place or manner of articulation, insert vowels, or delete consonants. Obstruent sounds, such as fricatives, affricates, and stops, can be especially difficult for L2 learners. This study analyzed native English speakers' perception of the quality of obstruents produced by native Mandarin Chinese and Korean speakers. Target words containing obstruents had been produced in three different tasks: in a carrier phrase, in a paragraph, and in a spontaneous speech sample. Obstruents were produced in word-initial position and word-final position. Raters with differing levels of expertise listened to these words and rated the perceptual quality of the obstruents within the words. This study found that overall, English obstruent productions by native Mandarin and Korean L2 speakers learning English were rated most clear when produced in word-initial position in a carrier phrase or a paragraph. The lowest ratings given were of obstruents in word-final position in spontaneous speech. No significant differences were found for listener expertise level. Combined with future research, results from this study will help educate the field of second language instruction as to how the speech of Korean and Mandarin learners of English is perceived. It also provides additional information on the effect that listener expertise has on the judgment of L2 speech production.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF STRUCTURE AND CONTENT

This thesis, *Perceptual Proficiency Ratings of Obstruent Productions in L2 Learners of English as a Function of Word Position, Task Type, and Listener Expertise,* was part of a larger collaborative project, portions of which may be submitted for publication, with the thesis author being one of multiple contributing coauthors. The body of this thesis was written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. The analyses conducted in this study were based on a set of recordings originally collected by Nissen, Hartshorn, Chase, Healy, Li, and Tanner (2016). An annotated bibliography is presented following the reference section in Appendix A. The consent form used in this study is found in Appendix B. Task type prompts are included in Appendix C.

Introduction

A number of studies (e.g., Baker, 2010; Boersma, 2010; Lee & Iverson, 2012) have examined the perceptual characteristics of English speech sounds in words produced by second language (L2) learners. However, these studies have examined this perception only when the words were produced in citation form and only when perceptual ratings were made by highly trained listeners. More knowledge on the effect of a listener's experience on their perceptual judgments of speech would be beneficial to the field of second language acquisition. The focus of the present study was the effect of word position on the perception of L2 English speech sounds, along with the study of words produced in other spoken contexts such as the oral reading of connected text and the telling of narratives.

Learning an L2 can be affected by multiple factors, such as the age of acquisition, number of years of previous linguistic experience, and the characteristics of an individual's native language (Baker, 2010; Hansen, 2001; Park & de Jong, 2008, Perani et al., 1998). A learner may need to learn entirely new phonological categories not found in their native language (Park & de Jong, 2008). One class of sounds that can be difficult for nonnative English speakers to acquire are obstruents, which include sounds such as stops, fricatives, and affricates. Obstruents are produced with a relatively constricted vocal tract. Stops, such as /p/, /t/, /k/, /b/, /d/ and /g/, are produced when air pressure builds up behind a completely obstruction in the vocal tract that is then opened, resulting in a rush of air from the constriction point (Davenport & Hannahs, 2005). Fricatives, such as /s/, /z/, /f/, /v/ and /θ/, are produced by turbulence or frication in the airflow as it moves through a relatively narrow constriction in the oral cavity (Jongman, Wayland, & Wong, 2000). Affricates, such as /ʧ/ and /ʤ/ are a combination of a stop

and a fricative that is produced when air is slowly released through a narrow opening in the oral cavity, as the tongue is gradually lowered (Davenport & Hannahs, 2005).

Studies have been conducted on the types of obstruents that are difficult for L2 learners of English to produce and identify, especially when a speech sound is not found in their native language (Baker, 2010; Park & de Jong, 2008). For example, the English fricatives /v/, /z/, /ʃ/, /ʒ/, /θ/, and /ð/ are not found in the Mandarin language. Second language learners of English frequently substitute the fricative /s/ for /θ/, and the fricative /f/ for /θ/ and for /v/ (Cheng, 1991; Slobin, 1992). Native speakers of Korean learning English as an L2 also have difficulty producing a number of English fricatives due in part to these sounds not being produced in their native language, such as /v/, /θ/, /ð/, /z/, /ʃ/, and /ʒ/. Therefore, when acquiring English, Korean L2 learners often substitute the fricative /θ/ with /s/ and the fricative /ð/ is often replaced with a stop, such as /d/ (Cheng, 1991).

Even if an obstruent speech sound is present in an L2 learner's native language, the speech sound may not occur in many linguistic contexts and word positions. In English, a wide variety of consonants frequently occur in word-final position resulting in closed syllables. In Mandarin the syllables are generally left open, with only the nasal stops /n/ and /ŋ/ used in word-final position (Fang & Ping-an, 1992; Nasukawa, 2004). Thus, it can be difficult for Mandarin L2 learners to produce obstruent speech sounds in word-final position. Hansen (2001) found that common errors made by Mandarin speakers when producing syllable-final codas in English included changing place or manner of articulation, inserting vowels, and deleting consonants. A study by He (2014) reported that Mandarin speakers learning English had relatively more difficulty producing consonants when in word-final position, with speakers often making deletion errors, changing features of sounds, and transferring properties of the Mandarin

language to English. Similar to Mandarin, the Korean language does not produce fricatives and affricate obstruents in word-final position. Also, sounds in the word-final position are usually devoiced in Korean and consonant clusters generally do not occur in word-initial or word-final positions (Cheng, 1991). These phonological and morphosyntactic characteristics of both Mandarin and Korean can interfere with an L2 learner's ability to accurately acquire and produce English obstruents.

The difficulty an L2 speaker may have in producing an English obstruent may also vary depending on the linguistic context or type of speaking task. Second language instructional methods often involve teaching new sound productions by having the learners repeat the targeted sound in isolation, single syllable words, or by practicing saying minimal pairs, such as "rip" and "lip", or "mouse" and "house." However, it is unclear how well L2 learners generalize their production of a speech sound to other types of more naturalistic speaking tasks involving longer utterances and spontaneous conversation. If the end goal of L2 instruction is to help the learner be an effective communicator in their new language, it is of interest to examine how well L2 pedagogy generalizes to more naturalistic contexts.

Previous studies have been conducted on the effects of task type on L2 language and speech performance. Tavakoli and Foster (2008) examined task type structure and recognized that it affected language complexity. It was found that when L2 learners of a variety of language backgrounds were given the assignment of narrating two cartoons, their performance was more complex when they were narrating the cartoon with a tightly structured story, as opposed to a loosely structured story. Skehan and Foster (1999) found that sequential and clearly structured tasks led to more fluent language production than less structured tasks. In contrast to the previously discussed findings, Gilabert, Barón, and Llanes (2009) observed that as a task

becomes more complex, L2 learners are more precise in their language and they work harder to complete the task successfully. Although several studies have explored the effect of task type on language performance (e.g., Skehan & Foster, 1999), research examining the effect of task type on speech and obstruent production is more limited.

Historically, obstruent production in native and nonnative speakers has been measured in a number of different ways. Some researchers have focused on measuring the acoustic and spectral components of obstruents. Forrest, Weismer, Milenkovic, and Dougall (1988) and Jongman et al. (2000) used spectral moment analysis to describe stop and fricative productions from typical English adult speakers in terms of the spectral mean, variance, skewness, and kurtosis. A study by Nissen and Fox (2005) examined these same spectral features in the fricative productions of younger children. Researchers have also focused on measuring the voice onset time of stop consonants. Lee and Iverson (2012) used voice onset time to analyze stops in Korean-English speaking children to evaluate how they develop phonetic systems. Obstruent production has also been measured using electropalatography, which tracks the linguapalatal contact patterns used to create a constriction in the vocal tract during the production of obstruent sounds (Fletcher, McCutcheon, & Wolf, 1975).

Perceptual ratings from native listeners are another commonly used method to evaluate obstruent productions (Boersma, 2010; Kreiman, Gerratt, & Precoda, 1990; Levi, Winters, & Pisoni, 2011). Perceptual ratings are beneficial in that an L2 learner's speech productions can be evaluated holistically by native listeners. The complexity of movement of each articulator, the transient nature of each sound, and loudness of a sound all affect a listener's ability to identify sounds. Hayes-Harb, Smith, Bent, and Bradlow (2008) studied the productions of voiceless word-final stop consonants in English by both English and native Mandarin speakers. These

productions were rated by a group of listeners. The value of using perceptual ratings was displayed in Yohiyuki and Brandenburger's study (1986) of fifth and sixth-grade elementary school students whose reading proficiency was rated by listeners. As more errors were made in the students' speech, their perceptual ratings dropped, and predictions of oral reading proficiency made after listening were found to be fairly accurate (Yohiyuki & Brandenburger, 1986). Listener ratings have also been used to examine speakers' vocal quality during speech production (Kreiman et al., 1990) and speech clarity (Munson, Johnson, & Edwards, 2012).

There are multiple ways that listeners can assign perceptual ratings to a speaker's speech productions (Boersma, 2010; Kreiman et al., 1990; Munson et al., 2012). Boersma had participants rate target sounds on a scale from one to four: one being very poor, two being fairly poor, three being fairly good, and four being very good. Kreiman's study had listeners rate voice quality for breathiness, roughness, abnormality, and instability by marking a point on a visual-analog scale for each category. Munson et al. had listeners assign perceptual ratings using a computerized visual analog scale. Raters selected a point between two target sounds on a scale to show to which target sound the production was closest, as well as to what degree it resembled that sound.

One question that needs to be addressed when using listener ratings, is what level of expertise regarding linguistics or communication sciences does a listener need to accurately rate an L2 learner's speech? In a study on perceptions of vocal quality, Kreiman et al. (1990) found that expert listeners relied on more variables than naïve listeners when making perceptual judgments on vocal quality. In another study on listener expertise, it was concluded that nonnative listeners comprehend nonnative speech better than native listeners (Hayes-Harb et al., 2008). Listener expertise was also examined in a study by Munson et al. (2012), in which the

differences between perceptual ratings done by speech-language pathologists (experienced listeners) and perceptual ratings conducted by individuals with no clinical experience (inexperienced listeners) were analyzed. All participants listened to young children's phoneme productions and were required to distinguish between the obstruent stop consonant pairs of /d/-/g/, /s/-/θ/, and /t/-/k/. The researchers found that the changes in the underlying properties of the speaker's sound productions influenced the experienced listeners' ratings differently than the inexperienced listeners' ratings. Overall, it was found that different levels of listening expertise had a significant influence on listeners' perceptual speech ratings and that the listening experience that individuals gain in a clinic aids them in being able to evaluate speech productions in a reliable manner when compared to listeners without specific training in linguistics or speech and language instruction.

Understanding how obstruent productions of L2 learners of English vary across differing speech tasks and word position may lead to improved pronunciation instruction in the classroom. It is also important to further examine how listeners' linguistic expertise may or may not affect their ability to accurately evaluate a speaker's L2 speech production. To further understand these issues, this study used listener ratings to examine how native Mandarin Chinese and Korean speakers' ability to produce English obstruents in an intelligible manner varied as a function of speech task type and word position. It was hypothesized that L2 learners would have more difficulty producing obstruents in an intelligible manner during more complex and less structured speech tasks, due in part to increased cognitive demands during these types of speech tasks. It was also hypothesized that L2 speakers would have more difficulty producing obstruents in word-final position, considering the relative lack of final consonant endings in the Korean and Mandarin languages. This study also sought to understand how listener expertise

might influence the perceptual assessment of L2 speech productions. It was hypothesized that the ratings of the L2 speech productions would change depending on the linguistic expertise level of the listeners.

Method

Participants

Participants for this study consisted of 10 undergraduate students and 10 graduate students enrolled in the Department of Communication Disorders at Brigham Young University, as well as eight expert evaluators experienced in the field of speech and language acquisition at Brigham Young University. Three separate types of listeners were evaluated: naïve listeners with little to no experience or instruction in evaluating speech, listeners who have taken basic courses in evaluating the phonetic target sounds in question, and expert listeners with experience in the assessment and treatment of L2 instruction. Each participant signed a consent form before being tested, and had a hearing screening via pure-tone air-conduction testing at 25 dB HL across one-octave intervals from 500-8000 Hz.

Stimuli

Speakers. Stimulus items for this study were extracted from audio recordings collected in a previous research project by Nissen, Hartshorn, Chase, Healy, Li, and Tanner (2016). The target stimuli in this study were produced by nine native speakers of Korean and twelve native speakers of Mandarin Chinese who were all learning English as a second language. The speakers were between 19 and 51 years of age, with an average age of 27. The speakers' English proficiency levels ranged from novice to advanced. On a proficiency rating scale from one to ten, the speakers ranged from three to eight. All ESL students involved in this study were

enrolled at Brigham Young University's English Language Center, a program designed to teach English to speakers of other languages.

Task type. The English words evaluated in this study contained the obstruents /f/, /θ/, /s/, /ʃ/, /p/ and /tʃ/ in the word-initial and word-final position. The words were elicited from the 21 L2 speakers across three different speech tasks. The three tasks included producing a set of target words in a carrier phrase and a paragraph reading, as well as eliciting spontaneous speech produced by explaining a culturally appropriate fairytale and answering open-ended questions in a conversational setting.

The stimulus words embedded in the carrier phrase *Say _____ again* were in monosyllabic form and structured in a CVC format. The target obstruent sounds were followed or preceded by a high front vowel (/i/ or /I/). The elicited words were *beef, feet, teeth, think, cheek, beach, leash, seat, sheep, piece, deep,* and *peek*. The speakers individually produced each carrier phrase three times, one of which was randomly selected for subsequent perceptual analysis.

In the paragraph speech task, the same words used as in the carrier phrases were embedded into two separate paragraphs, as listed in Appendix C. Thus, each target stimulus word was read twice, of which one instance was randomly selected for evaluation by the native English listeners.

The spontaneous speech samples were elicited by having speakers recall a culturally appropriate fairytale and answer a series of open-ended questions, as listed in Appendix C. Due to the spontaneous nature of these samples, the target sounds preceded or were followed by a variety of English vowels. One of each stimulus word from the spontaneous speech portion that contained the target obstruent in either word-initial or word-final position was randomly

selected. Two hundred and sixty-six words were extracted from the spontaneous speech samples. Some of the target sounds in certain word positions were not elicited in the spontaneous speech task by all speakers either due to participant's evasion of the sound or a lack of commonly produced words with the target sound in either the word-initial or word-final position.

Recordings. The audio recordings of the stimulus items were collected in a sound-attenuating booth using a high-quality, low-impedance dynamic microphone and preamplifier. Recordings were made with a sampling rate of 44.1 kHz and a quantization of 24 bits. All of the stimulus words were cut from the raw recordings by selecting the beginning and end of each word using Adobe Audition (San Jose, CA) sound editing software. The stimulus words were normalized for intensity and filtered for electronic noise and noise artifacts. The duration of the clipped speech recordings was edited to have 500 ms of silence preceding and following each word.

Procedures

Each participant completed the listener ratings in two sessions, with each session lasting for forty-minutes. The data collection was divided into four stages: an introduction including a hearing screening and instructions, and three 20-minute testing periods in which participants listened to and rated the stimulus words. Each 20-minute period was divided by a three-minute break.

The stimulus words were grouped according to the specific obstruent being targeted for evaluation and randomly presented to the listeners. All audio signals were presented via headphones to participants who could then adjust the stimulus volume from a starting level of 60 dB HL to a comfortable level. Volume was kept within the range of safe hearing. Each

participant evaluated a practice trial of each stimulus word type to ensure they understood the rating system, and the equipment was adjusted properly. Participants were instructed to listen for a specific phoneme (/f/, /θ/, /s/, /ʃ/, /p/, or /tʃ/) at the beginning or end of each word and then rate their perception of the clarity of the obstruent production using a sliding analog scale, similar to the scale used in the study by Munson et al. (2012). The scale was labeled with 0 on one end and 100 on the other end (0 corresponding to a completely distorted production and 100 corresponding to a typical, undistorted production), with a tick mark in the middle of the scale to indicate a rating of 50. After rating a production, participants were instructed to advance to the next stimulus item by using the mouse to select a small box on a computer screen.

Intra-rater and Listener Group Reliability

To examine the intra-rater reliability within a listener's ratings, 10 percent of the stimulus items were rated a second time per listener. Pearson Correlations was used in order to determine intra-rater reliability and differences between listener groups. Overall, the first and second sets of listener ratings had a correlation of $r = .77$, $p < 0.001$. The individual listener's ratings had correlations between a low of $r = .54$ and a high of $r = .84$. The average correlation was $r = .73$ for the inexperienced listener group, $r = .74$ for the experienced listener group, and $r = .71$ for the expert listener group.

Statistical Analysis

A mixed-model repeated measures analysis of variance (ANOVA) was performed to analyze these data with a between-subjects factor of listener expertise and the within-subject factors of task type and word position. The 10 percent of stimulus items that were rated a second time for intra-rater reliability were not included in the ANOVA. The ANOVA results include a measure of effect size, partial eta squared, or partial η^2. The value of partial η^2 can range from

0.0 to 1.0, and is considered a proportion of variance explained by a dependent variable when controlling for other factors.

Results

This study was part of a larger project designed to evaluate the pronunciation abilities of native Mandarin and Korean L2 learners of English conducted by Nissen et al. (2016).

Mandarin L2 Productions

Listener expertise. Intermediate and expert listeners differed significantly in their ratings of /f/. Expert listeners rated the /f/ sound produced by Mandarin speakers as less articulate ($M = 63.29$, $SD = 3.15$) than the intermediate raters ($M = 76.24$, $SD = 2.76$), $F(2, 78) = 5.10$, $p < .05$, partial $\eta^2 = .12$. Naïve listeners rated more similarly to intermediate listeners when rating /f/ ($M = 73.41$, $SD = 2.86$). There were no significant differences across listener experience among /θ/, /s/, /ʃ/, /p/, and /tʃ/ produced by Mandarin speakers.

Task type and word position. The ratings of the voiceless postalveolar affricate /tʃ/ produced by Mandarin speakers differed significantly across task type, $F(2, 160) = 22.28$, $p < .001$, partial $\eta^2 = .22$. The affricate was rated most articulate when produced in a paragraph ($M = 81.48$, $SD = 1.66$) and least articulate in spontaneous speech ($M = 67.68$, $SD = 2.14$). Ratings of /tʃ/ also differed significantly across task type and word position, $F(2, 160) = 17.39$, $p < .001$, partial $\eta^2 = .18$. As shown in Figure 1, /tʃ/ was rated most clear in the word-initial position in paragraphs ($M = 85.64$, $SD = 2.04$). It was rated least clear when produced in the word-initial position in spontaneous speech ($M = 61.56$, $SD = 3.32$). No other comparisons were found to differ significantly. Please refer to Table 1 for all of the remaining descriptive statistics associated with Mandarin speaker productions.
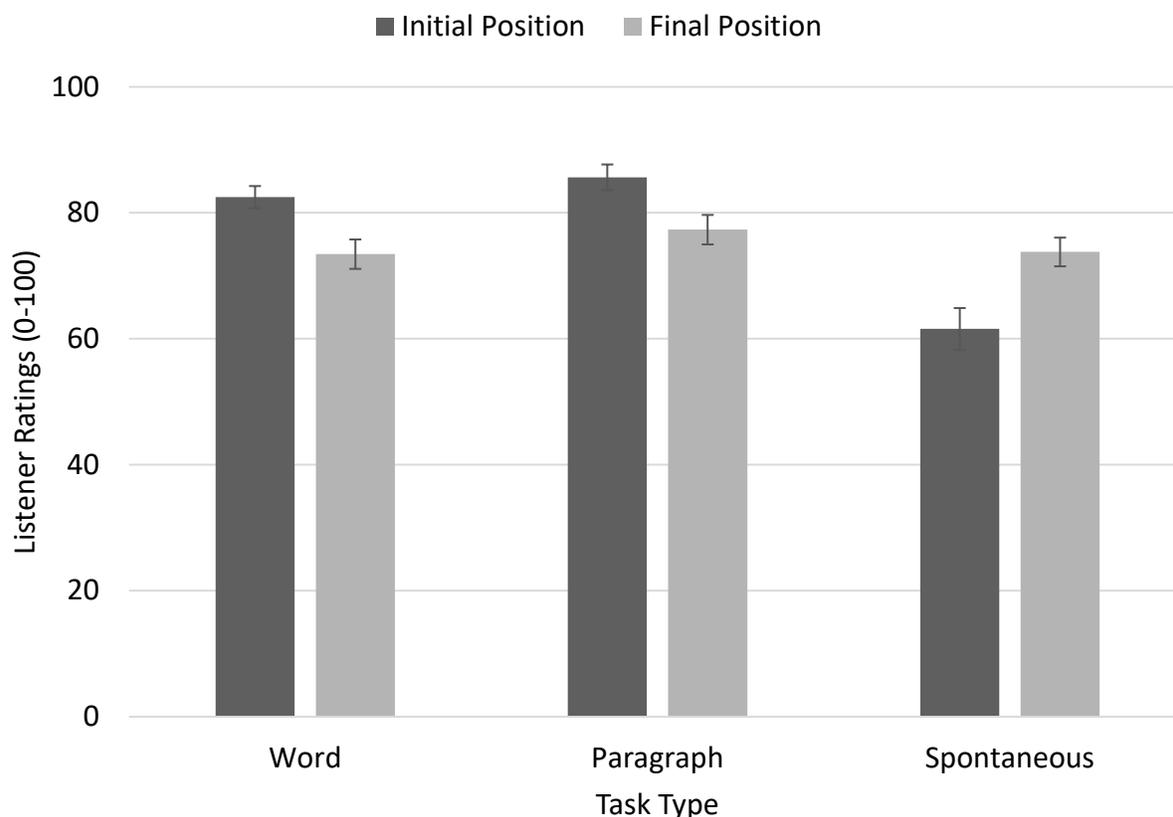
Figure 1. Listener ratings for Mandarin speaker consonant productions of /tʃ/ across task type and word position.

The ratings of the voiceless labiodental fricative /f/ produced by Mandarin speakers differed significantly across word position, $F(1, 80) = 91.88$, $p < .001$, partial $\eta^2 = .54$. The fricative in word-initial position was rated more articulate ($M = 80.12$, $SD = 1.89$) than word-final position ($M = 63.05$, $SD = 2.06$). No other comparisons differed significantly when comparing task type.

The ratings of the voiceless bilabial stop /p/ produced by Mandarin speakers differed significantly across task type, $F(2, 160) = 79.35$, $p < .001$, partial $\eta^2 = .50$. The consonant was rated highest when produced in paragraphs ($M = 84.90$, $SD = 1.82$) and lowest in spontaneous speech ($M = 56.14$, $SD = 2.60$). The ratings of the stop differed significantly across word
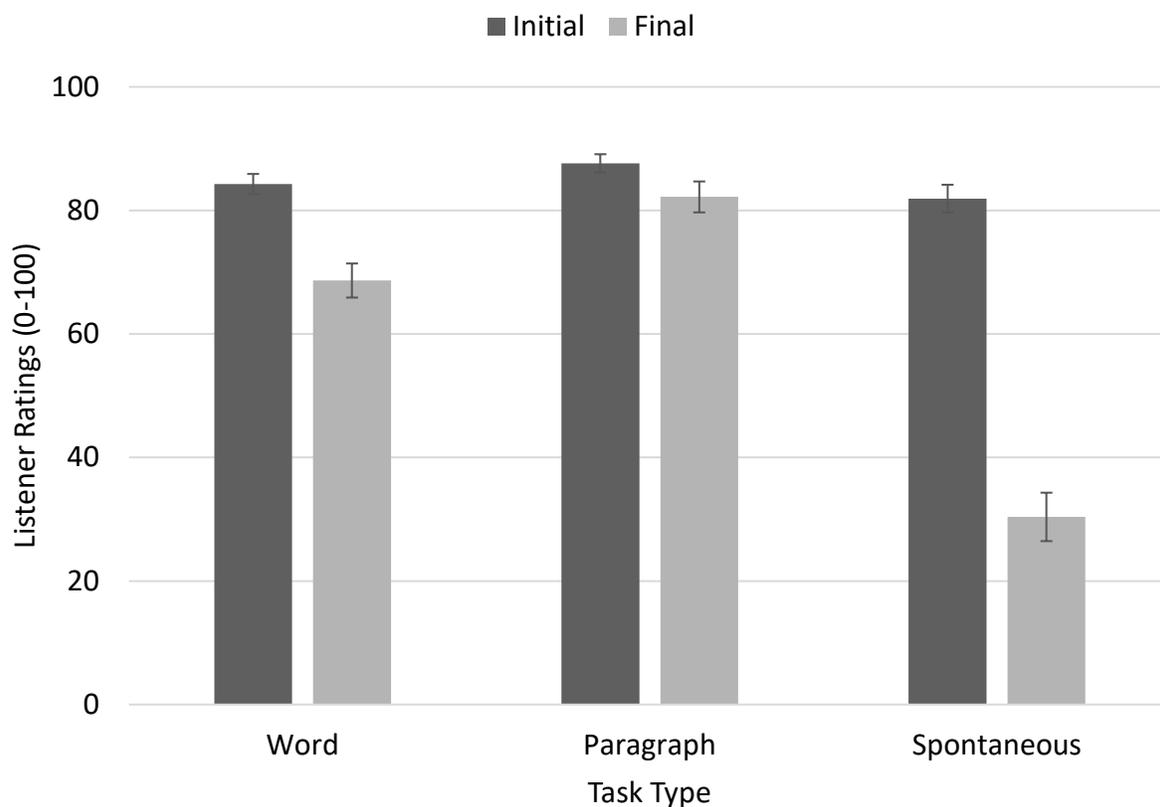
Figure 2. Listener ratings for Mandarin speaker consonant productions of /p/ across task type and word position.

position, $F(1, 80) = 195.13$, $p < .001$, partial $\eta^2 = .71$. The consonant in word-final position ($M = 60.40$, $SD = 2.16$) had lower ratings than in word-initial position ($M = 84.60$, $SD = 1.50$). The ratings of /p/ produced by Mandarin speakers differed significantly across task type and word position, $F(2, 160) = 78.57$, $p < .001$, partial $\eta^2 = .50$. As shown in Figure 2, /p/ was rated highest when produced in word-initial position in a paragraph ($M = 87.62$, $SD = 1.47$). It was rated lowest when produced in word-final position in spontaneous speech ($M = 30.38$, $SD = 3.92$). No other comparisons differed significantly.

The ratings of the voiceless alveolar fricative /s/ produced by Mandarin speakers differed significantly across task type, $F(2, 160) = 12.82$, $p < .001$, partial $\eta^2 = .14$. The fricative was

Table 1

Listener ratings of Mandarin speaker consonant productions across task type and word position

Sound	Task Type	Initial Position		Final Position	
		M	*SD*	*M*	*SD*
Postalveolar affricate /tʃ/	Word	82.49	1.77	73.44	2.34
	Paragraph	85.64	2.04	77.33	2.34
	Spontaneous	61.56	3.32	73.79	2.28
Labiodental fricative /f/	Word	78.72	2.63	69.01	2.73
	Paragraph	81.20	2.34	59.99	3.15
	Spontaneous	80.45	2.46	60.17	3.61
Bilabial stop /p/	Word	84.27	1.64	68.65	2.76
	Paragraph	87.63	1.47	82.18	2.51
	Spontaneous	81.90	2.25	30.38	3.92
Alveolar fricative /s/	Word	74.42	2.75	67.45	2.97
	Paragraph	61.41	3.64	61.23	3.04
	Spontaneous	69.28	2.70	53.58	3.34
Postalveolar fricative /ʃ/	Word	80.30	2.23	57.79	3.35
	Paragraph	79.90	2.23	78.41	2.49
	Spontaneous	71.31	2.41	55.05	3.33
Interdental fricative /θ/	Word	80.13	2.25	65.29	3.29
	Paragraph	80.29	2.42	57.23	3.54
	Spontaneous	37.82	3.84	29.05	3.88

rated most articulate when produced in a carrier phrase ($M = 70.94$, $SD = 2.39$), while the

paragraph ($M = 61.32$, $SD = 2.88$) and spontaneous speech ($M = 61.43$, $SD = 2.51$) ratings were

similar. The /s/ ratings also differed significantly across word position, $F(1, 80) = 12$, $p < .05$,

partial $\eta^2 = .13$. The fricative /s/ in word-initial position ($M = 68.37$, $SD = 2.40$) had higher

ratings than in word-final position ($M = 60.75$, $SD = 2.65$). In addition, the listener ratings of /s/

differed significantly as a function of task type and word position, $F(2, 160) = 6.41$, $p < .05$,

partial $\eta^2 = .07$. The /s/ was rated most articulate when produced in the word-initial position in a

carrier phrase ($M = 74.42$, $SD = 2.75$). The /s/ was rated least articulate when produced in word-

final position in spontaneous speech ($M = 53.58$, $SD = 3.34$). As shown in Figure 3, the /s/

Figure 3. Listener ratings for Mandarin speaker productions of /s/ across task type and word

position.

elicited in a paragraph was rated similarly in both word positions, but ratings differed between word-initial and word-final position within spontaneous speech and within a word in a carrier phrase. No other comparisons were found to differ significantly for the /s/ sound.
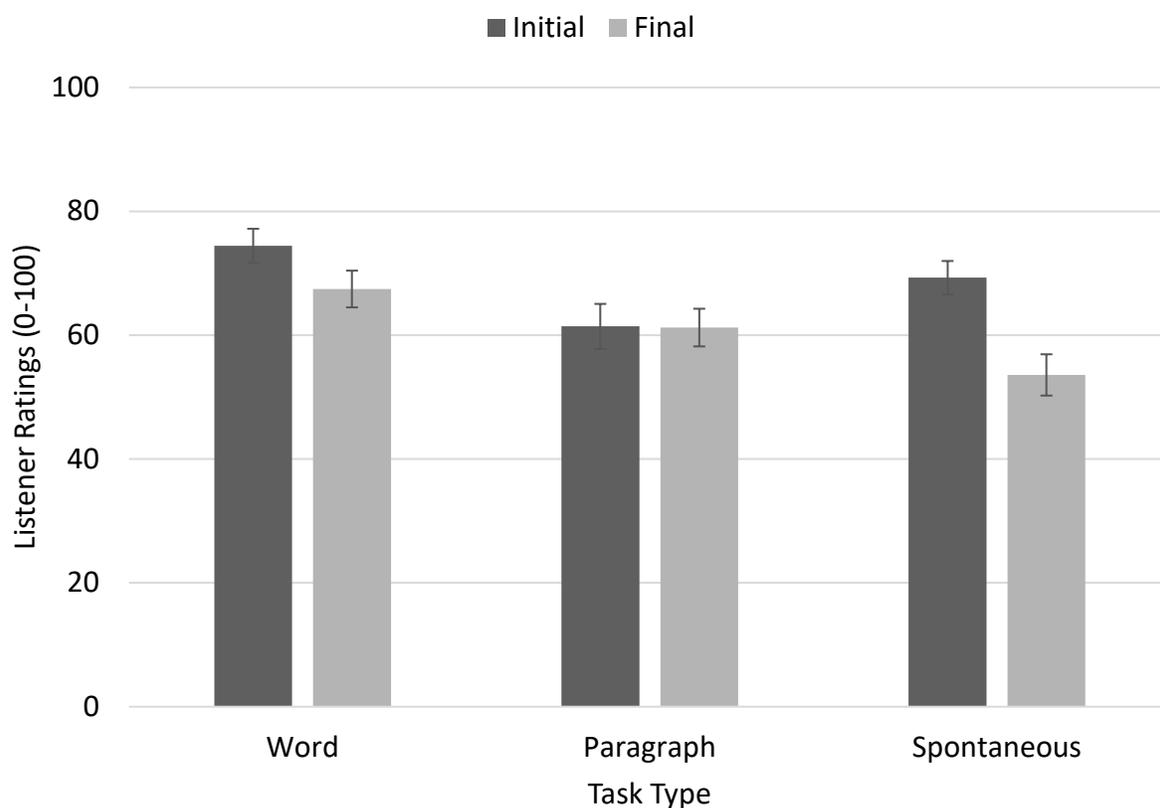
The ratings of the voiceless postalveolar fricative /ʃ/ produced by Mandarin speakers differed significantly across task type, $F(2, 160) = 33.50$, $p < .001$, partial $\eta^2 = .30$. The fricative was rated highest when produced in a paragraph ($M = 79.18$, $SD = 1.97$) and lowest in spontaneous speech ($M = 63.18$, $SD = 2.25$). The ratings of the fricative /ʃ/ differed significantly across word position, $F(1, 80) = 46.65$, $p < .001$, partial $\eta^2 = .37$. The consonant in word-final position ($M = 63.75$, $SD = 2.31$) had lower ratings than in word-initial position ($M = 77.18$, $SD = 1.83$). The ratings of the fricative /ʃ/ produced by Mandarin speakers also differed significantly across task type and word position, $F(2, 160) = 11.25$, $p < .001$, partial $\eta^2 = .12$. As shown in Figure 4, the /ʃ/ sound was rated highest when produced in word-initial position in a carrier phrase ($M = 80.30$, $SD = 2.23$). It was rated lowest when produced in word-final position in spontaneous speech ($M = 55.05$, $SD = 3.33$). No other comparisons were found to differ significantly.

The ratings of the voiceless interdental fricative /θ/ produced by Mandarin speakers differed significantly across task type, $F(2, 160) = 132.11$, $p < .001$, partial $\eta^2 = .62$. The fricative was rated most articulate when produced in a carrier phrase ($M = 72.71$, $SD = 2.28$) and lowest in spontaneous speech ($M = 33.43$, $SD = 2.50$). The ratings for the interdental fricative /θ/ produced by Mandarin speakers differed significantly across word position, $F(1, 80) = 35.08$, $p < .001$, partial $\eta^2 = .31$. The consonant in word-final position ($M = 50.52$, $SD = 2.36$) had lower ratings than in word-initial position ($M = 66.08$, $SD = 2.13$). No other comparisons differed significantly.
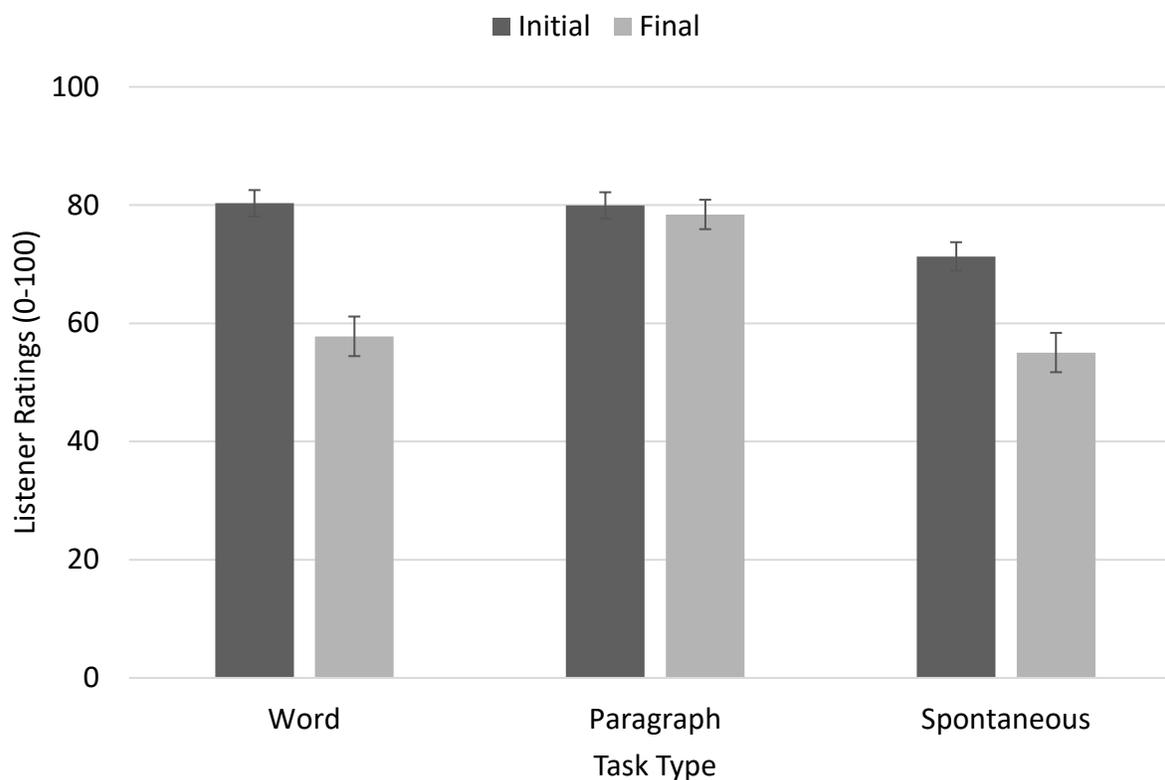
Figure 4. Listener ratings for Mandarin speaker productions of /ʃ/ across task type and word position.

Korean L2 Productions

Listener expertise. There was no significant difference between the ratings across listener experience for stimuli of the Korean speakers.

Task type and word position. The ratings of the voiceless postalveolar affricate /tʃ/ produced by Korean speakers differed significantly across task type, $F(2, 54) = 15.41, p < .001$, partial $\eta^2 = .36$. The affricate was rated most articulate when produced in a paragraph ($M = 79.36, SD = 2.73$) and least articulate in spontaneous speech ($M = 59.57, SD = 4.43$). The ratings of the affricate /tʃ/ produced by Korean speakers also differed significantly across word position, $F(1, 27) = 15.42, p < .05$, partial $\eta^2 = .36$. The affricate was rated most articulate in word-initial position ($M = 77.88, SD = 2.77$) when compared to word-final position ($M = 65.49, SD = 3.41$).

There were no other significant differences for /tʃ/. Please refer to Table 2 for all of the remaining descriptive statistics associated with Mandarin speakers productions.

The ratings of the voiceless labiodental fricative /f/ produced by Korean speakers differed significantly across task type, $F(2, 54) = 34.56$, $p < .001$, partial $\eta^2 = .56$. The fricative was rated highest when produced in a paragraph ($M = 83.21$, $SD = 3.04$) and lowest in spontaneous speech ($M = 54.70$, $SD = 2.77$). The ratings of the fricative /f/ produced by Korean speakers differed significantly across task type and word position, $F(2, 54) = 204.97$, $p < .001$, partial $\eta^2 = .88$. As shown in Figure 5, the /f/ was rated highest when produced in word-initial position in spontaneous speech ($M = 91.52$, $SD = 2.23$). It was rated lowest when produced in word-final position in spontaneous speech ($M = 17.89$, $SD = 4.48$). No other comparisons differed significantly.

The ratings of the voiceless bilabial stop /p/ produced by Korean speakers were significantly different across task type, $F(2, 54) = 110.31$, $p < .001$, partial $\eta^2 = .80$. The /p/ sound was rated highest when produced in a carrier phrase ($M = 79.28$, $SD = 3.40$) and lowest in spontaneous speech ($M = 36.31$, $SD = 3.11$). The ratings of the stop /p/ produced by Korean speakers were also significantly different across word position, $F(1, 27) = 30.80$, $p < .001$, partial $\eta^2 = .53$. The /p/ was rated most articulate in word-initial position ($M = 73.24$, $SD = 3.60$) when compared to word-final position ($M = 56.25$, $SD = 2.44$). The ratings of the voiceless bilabial stop /p/ produced by Korean speakers differed significantly across task type and word position, $F(2, 54) = 41.47$, $p < .001$, partial $\eta^2 = .61$. As shown in Figure 6, /p/ was rated highest in word-initial position in a carrier phrase ($M = 80.80$, $SD = 3.92$). The /p/ was rated lowest in word-final position in spontaneous speech ($M = 9.78$, $SD = 2.38$). No other comparisons were significantly different.

Table 2

Listener ratings of Korean speaker consonant productions across task type and word position

Sound	Task Type	Initial Position		Final Position	
		M	*SD*	*M*	*SD*
Postalveolar affricate /tʃ/	Word	83.71	3.11	68.53	4.34
	Paragraph	82.98	2.62	75.76	3.86
	Spontaneous	66.94	5.64	52.19	5.54
Labiodental fricative /f/	Word	74.69	4.12	83.13	3.11
	Paragraph	77.89	4.10	88.54	2.29
	Spontaneous	91.52	2.23	17.89	4.48
Bilabial stop /p/	Word	80.80	3.92	77.76	4.14
	Paragraph	76.06	4.11	81.22	4.37
	Spontaneous	62.85	5.35	9.78	2.38
Alveolar fricative /s/	Word	21.29	5.55	66.51	5.30
	Paragraph	72.43	4.71	58.94	5.18
	Spontaneous	71.50	4.72	72.63	4.68
Postalveolar fricative /ʃ/	Word	84.60	3.18	79.26	4.38
	Paragraph	79.46	3.53	85.69	3.16
	Spontaneous	60.27	5.24	15.57	3.52
Interdental fricative /θ/	Word	88.60	2.24	75.46	3.92
	Paragraph	79.29	3.90	83.33	3.25
	Spontaneous	57.57	5.68	25.18	5.13
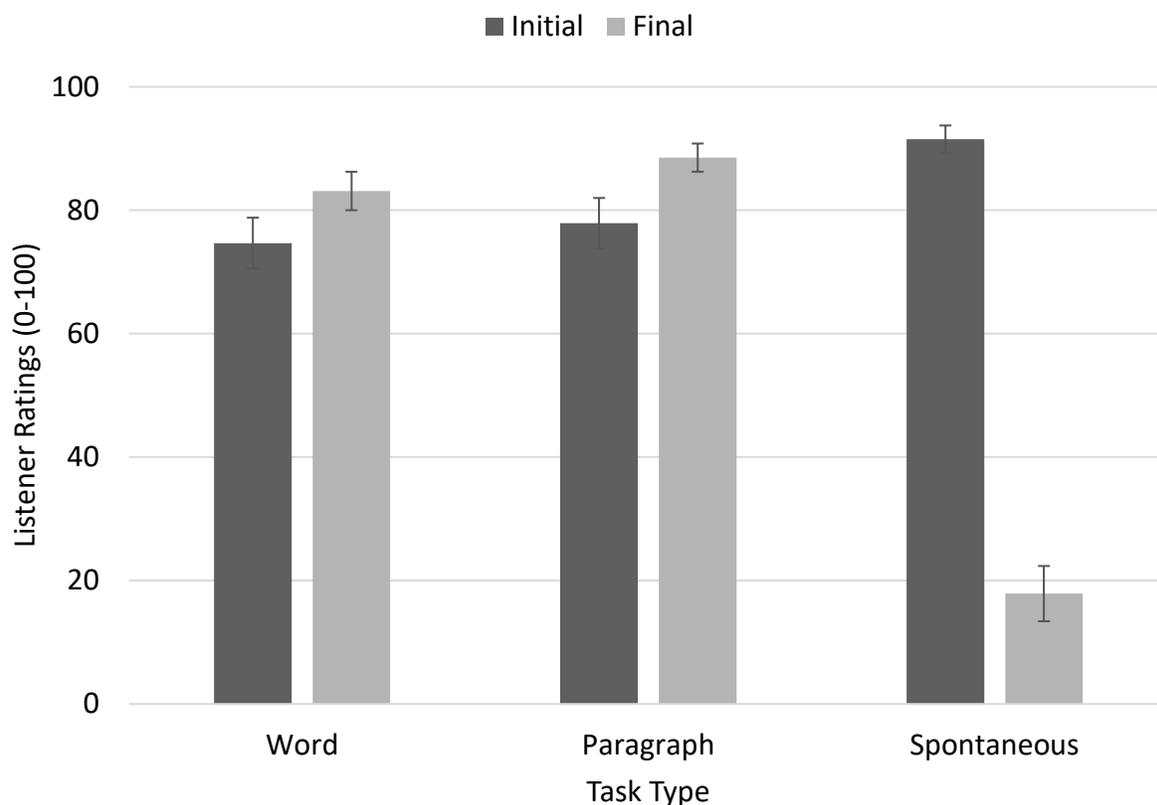
Figure 5. Listener ratings for Korean speaker consonant productions of /f/ across task type and word position.

The ratings of the voiceless alveolar fricative /s/ produced by Korean speakers differed significantly across task type, $F(2, 54) = 36.12$, $p < .001$, partial $\eta^2 = .57$. The fricative was rated highest when produced in spontaneous speech ($M = 72.06$, $SD = 4.14$) and lowest in a carrier phrase ($M = 43.90$, $SD = 4.12$). The ratings of the fricative /s/ produced by Korean speakers differed significantly across word position, $F(1, 27) = 8.99$, $p < .05$, partial $\eta^2 = .25$. The affricate was rated most articulate when produced in word-final position ($M = 66.03$, $SD = 4.08$) when compared to the word-initial position ($M = 55.07$, $SD = 4.05$). The ratings of the voiceless alveolar fricative /s/ produced by Korean speakers differed significantly across task type and word position, $F(2, 54) = 32.41$, $p < .001$, partial $\eta^2 = .55$. As shown in Figure 7, /s/ was rated

Figure 6. Listener ratings for Korean speaker consonant productions of /p/ across task type and word position.

highest when produced in word-final position in spontaneous speech (*M* = 72.63, *SD* = 4.68). It was rated lowest when produced in the word-initial position in a carrier phrase (*M* = 21.29, *SD* = 5.55). No other comparisons differed significantly.

The ratings of the voiceless postalveolar fricative /ʃ/ produced by Korean speakers were significantly different across task type, *F*(2, 54) = 106.51, *p* < .001, partial η^2 = .80. The fricative was rated highest when produced in a paragraph (*M* = 82.58, *SD* = 3.14) and lowest in spontaneous speech (*M* = 37.92, *SD* = 3.55). The ratings of the fricative /ʃ/ produced by Korean speakers differed significantly across word position, *F*(1, 27) = 42.26, *p* < .001, partial η^2 = .61. The fricative was rated most articulate when produced in word-initial position (*M* = 74.78, *SD* =
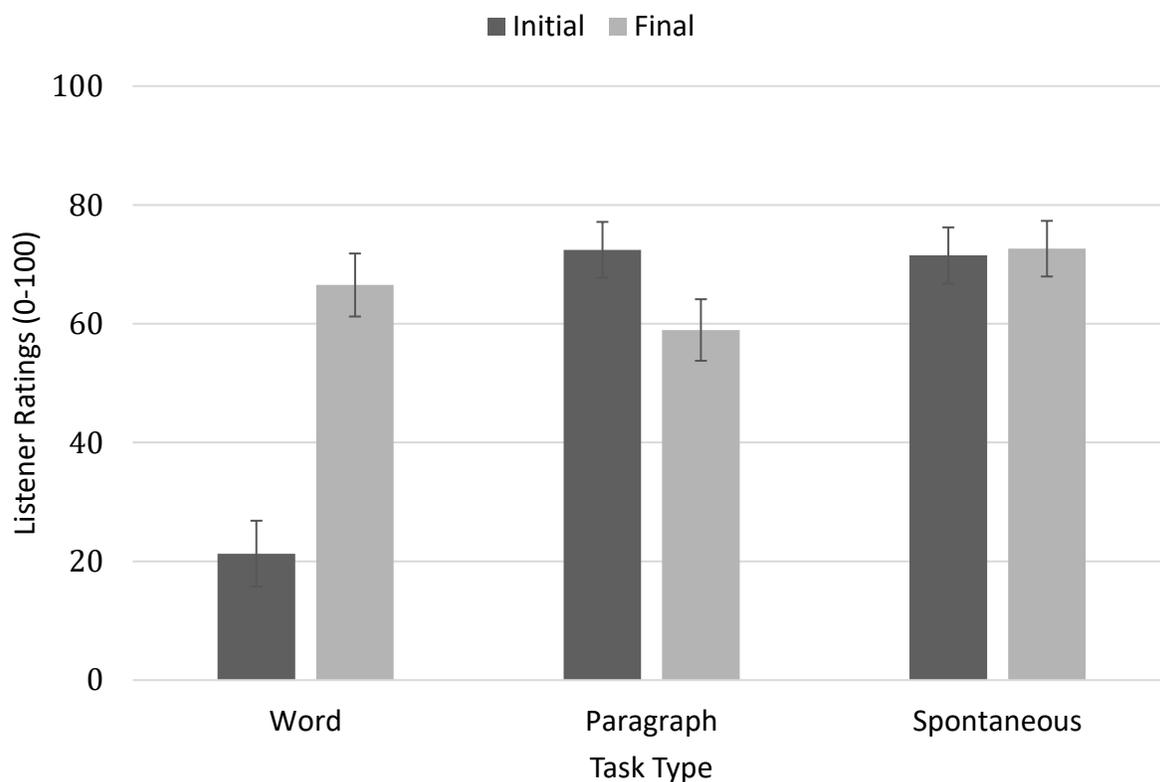
Figure 7. Listener ratings for Korean speaker consonant productions of /s/ across task type and word position.

2.85) when compared to the word-final position ($M = 60.17$, $SD = 2.53$). The ratings of the voiceless postalveolar fricative /ʃ/ produced by Korean speakers differed significantly across task type and word position, $F(2, 54) = 30.27$, $p < .001$, partial $\eta^2 = .53$. As shown in Figure 8, /ʃ/ was rated highest when produced in word-final position in a paragraph ($M = 85.69$, $SD = 3.16$). It was rated lowest when produced in word-final position in spontaneous speech ($M = 15.57$, $SD = 3.52$). No other comparisons differed significantly.

The ratings of the voiceless interdental fricative /θ/ produced by Korean speakers differed significantly across task type, $F(2, 54) = 85.53$, $p < .001$, partial $\eta^2 = .76$. The fricative was rated highest when produced in a carrier phrase ($M = 82.03$, $SD = 2.77$) and lowest in spontaneous speech ($M = 41.38$, $SD = 4.50$). The ratings of the interdental fricative /θ/ produced by Korean
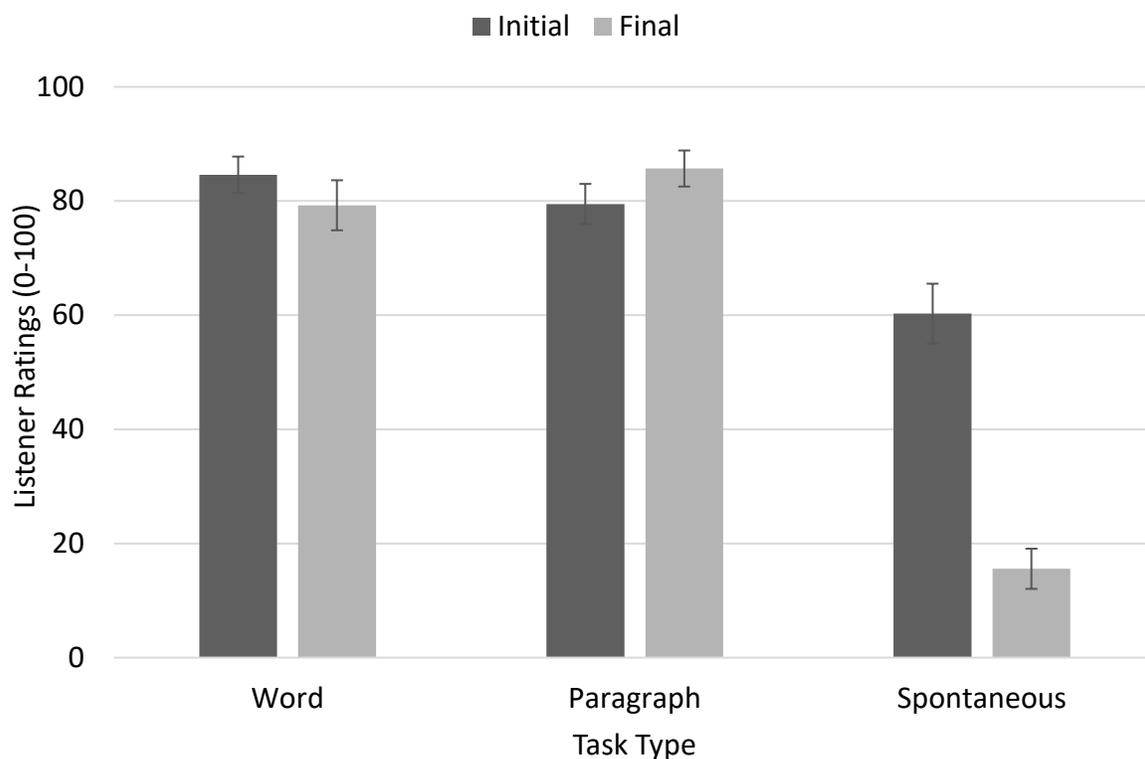
Figure 8. Listener ratings for Korean speaker consonant productions of /ʃ/ across task type and word position.

speakers differed significantly across word position, $F(1, 27) = 38.99$, $p < .001$, partial $\eta^2 = .59$. The fricative was rated highest when produced in word-initial position ($M = 75.15$, $SD = 2.92$) when compared to word-final position ($M = 61.33$, $SD = 3.21$). The ratings of the voiceless interdental fricative /θ/ produced by Korean speakers were significantly different across task type and word position, $F(2, 54) = 14.55$, $p < .001$,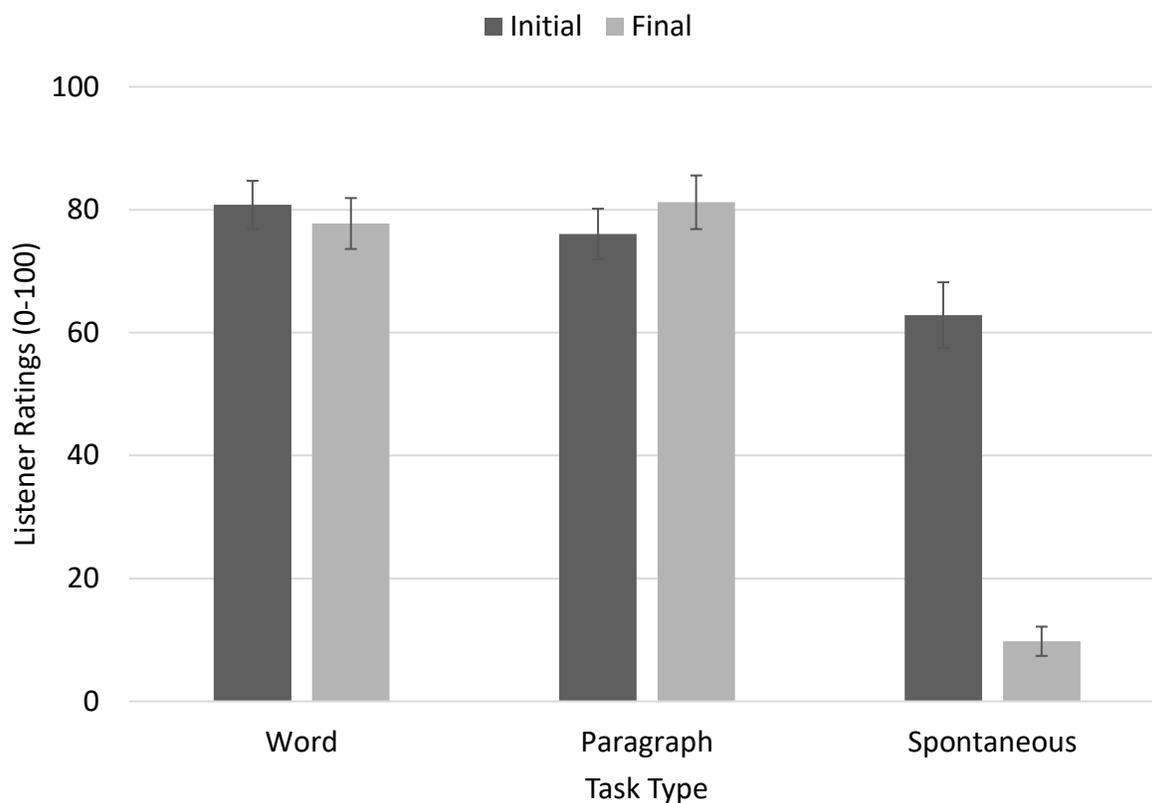 partial $\eta^2 = .35$. As shown in Figure 9, /θ/ was rated highest when produced in word-initial position in a carrier phrase ($M = 88.60$, $SD = 2.24$). It was rated lowest when produced in word-final position in spontaneous speech ($M = 25.18$, $SD = 5.13$). No other comparisons differed significantly.
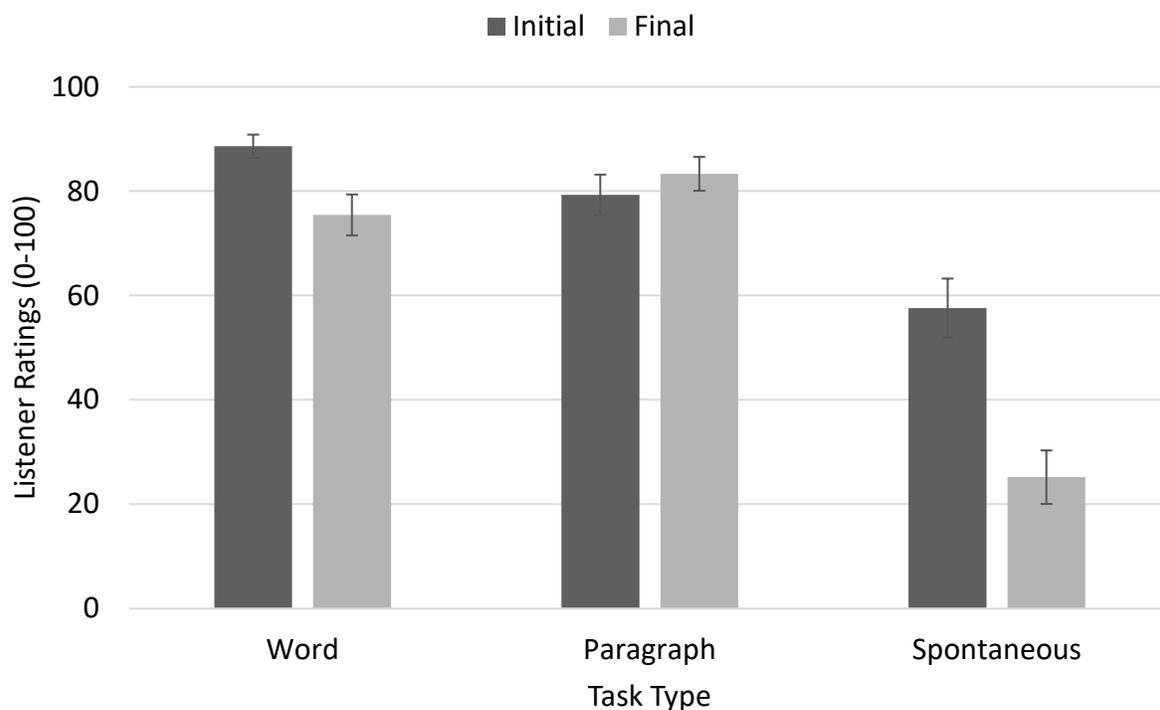
Figure 9. Listener ratings for Korean speaker consonant productions of /θ/ across task type and word position.

Discussion

One hypothesis of this study was that native English listeners would rate L2 learners' obstruent productions during more structured speech tasks with a higher degree of pronunciation clarity. This hypothesis was based in part on the expectation that a more structured task would require less cognitive resources to formulate and thereby result in more clear speech production. Previous findings by Tavakoli and Foster (2008) and Skehan and Foster (1999), found that speakers produced more fluent language when given tightly structured language tasks. These findings are in contrast to the conclusions of Gilabert, Barón, and Llanes (2009), who observed that as a task becomes more complex, L2 learners are more precise in their speech communication.

The results of this study were more similar to the findings of Skehan and Foster (1999) and Tavakoli and Foster (2008), in that the L2 speakers' obstruent productions were generally found to have more clarity when elicited in the carrier phrase or paragraph conditions, compared to the spontaneous speech condition. Mandarin speakers' obstruent productions on average were rated lowest in pronunciation clarity when elicited in a spontaneous speech task, with exception of the voiceless labiodental fricative /f/ productions. Results from the Korean L2 speakers were more varied across obstruent sound type. The /p/ and /θ/ speech sounds were rated as having the greatest clarity when produced in carrier phrases, whereas listeners found /tʃ/ and /ʃ/ sounds to be more clearly produced when the speaker was reading a paragraph. Unlike the other speech sounds, the Korean speakers' production of the /f/ and /s/ sounds were found to have the highest pronunciation clarity when spoken in the spontaneous speech tasks.

It is clear that L2 learners experience some level of difficulty generalizing productions of English obstruents from citation form to more complex tasks, like spontaneous speech. This is a cause for concern in a student's language learning experience, as the final goal for a student is possessing functional speech and language skills that are more complex in nature. Second language instructors assessing a student's language abilities would benefit from implementing a comprehensive assessment of language and speech that includes more complex tasks, like conversational speech and narratives, along with simpler tasks of reading aloud and providing short answers. It would also be beneficial for L2 instructors to perceptually evaluate a learner's speech periodically, using narratives and conversation topics as prompts.

Also hypothesized in this study was that listeners would rate the L2 learners' productions in word-final position as having less pronunciation clarity than those produced in the word-initial position. This hypothesis was based on the relative lack of final consonant endings in both

Mandarin and Korean (Cheng, 1991; Fang & Ping-an, 1992; Nasukawa, 2004). This study generally found that the L2 learners' obstruent productions were rated to have greater pronunciation clarity when produced in the word-initial position compared to word-final position. Two of the phonemes that are not found in Mandarin (/ʧ/ and /θ/) were found to have the lowest clarity when in the word-final position. The lowest ratings for clarity were associated with word-final obstruent productions in the spontaneous speech task. Likewise, listeners found the least degree of pronunciation clarity for the obstruent sounds in word-final position that are not found in the Korean language, such as /f/, /θ/, and /ʃ/. According to this study, L2 learners who speak native languages lacking certain final consonant sounds are judged by others as having less clarity in final consonants productions in the L2. Second language instructors would benefit from knowing the qualities of an L2 learner's native language in order to discern what sounds may be most difficult because they do not appear in the native language or do not occur in certain positions. They could then plan instruction more focused on improving these sounds.

Results of this study found that listeners rated Mandarin speakers' productions of /tʃ/, /f/, /p/, /s/, /ʃ/, and /θ/ to be more intelligible in word-initial position across all task types and less intelligible in spontaneous speech in word-final position. Exceptions were noted for /tʃ/ in spontaneous speech, while /f/ was not significantly different for task type. Listeners judged all Korean speakers' productions of /tʃ/, /f/, /p/, /ʃ/, and /θ/ to be more intelligible in word-initial position, with the exception of the fricative /s/. The results of this study aligned with the findings of He (2014) and Hansen (2001) who found that Mandarin speakers have difficulty producing final consonants in English. In the studies by He and Hansen, speakers made a variety of errors that included the change of place or manner of articulation, vowel insertion, first language (L1) interference, and consonant deletion. This difficulty with word-final consonants

and the presence of these errors is evident to us through significantly lower ratings of word-final productions.

It was hypothesized that the ratings of expert, experienced, and inexperienced listeners would be significantly different, due to differing levels of knowledge in the field of L2 acquisition. The majority of statistical tests were not found to be statistically significant for listener expertise. This finding contradicts the conclusions of Munson et al. (2012) who found that listener expertise had a significant influence on perceptual ratings of speech productions. Students learning English receive specific L2 instruction in an L2 classroom, but they also speak outside the classroom, with their speech being evaluated by many people in their entire community. This study concludes that despite a lack of extensive experience in judging speech production or teaching English as a second language, experienced and inexperienced listeners' ratings are similar to those of expert listeners, in that they rate the level of clarity of English learners' speech productions similarly. Based on Yohiyuki and Brandenburger's (1986) findings that listeners are reliable at giving lower perceptual ratings as errors increase, and that perceptual ratings can be used to predict reading proficiency, this current study can help us to know that perceptual judgments made by listeners of all expertise levels may be useful towards predicting a student's reading proficiency.

One limitation of the current study is the lack of an ideal pronunciation standard from which the listeners could base their ratings. In future research, it may be valuable to compare perceptual ratings of L1 English obstruent productions with the ratings of L2 English obstruent productions. This will allow an analysis of what is ideal in English speech, and how clear native English speech is compared to different levels of L2 English speech. Another recommendation is to have a greater number of raters in each listener expertise group. This study was also limited

in that some of the L2 speakers would avoid producing words that contained certain sounds in the spontaneous speech task. In future research, participants could be given a list of words from which they construct a spontaneous story, thereby increasing the chance that each target sound would be produced in both word-initial and word-final position. Pictures could also be used to represent the target words to avoid the possible influence of a speaker's reading ability.

Despite these limitations, this study provides additional information concerning how perceptual ratings of native Mandarin and Korean L2 learners' production of English obstruents may change as a function of linguistic task type and word position of a sound, and whether the experience level of a listener has a significant impact on their evaluation of L2 pronunciation. These findings may lead to improved assessment and instruction in the classroom for L2 learners of English. This study is unique in that it uses listeners with varying levels of listener expertise to perceptually evaluate obstruent productions in a variety of task types produced by native Mandarin and Korean L2 learners of English, an area of research that is underdeveloped within the field of L2 acquisition.

References

Adobe Systems Incorporated. (2013). Adobe Audition CC (Version 6) [Computer software]. San Jose, CA: Adobe Systems Incorporated.

Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition*, *13*, 263-278. doi:10.1017/S136672890999006X

Boersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *Journal of the Acoustical Society of America*, *127,* 1636-1644. doi:10.1121/1.3292996

Cheng, L. (1991). *Assessing Asian language performance: Guidelines for evaluating limited-English proficient students* (2nd ed.). Oceanside, CA: Academic Communication Associates.

Davenport, M., & Hannahs, S., (2005). *Introducing phonetics & phonology*. London, UK. Hodder Education.

Fang, X., & Ping-an, H. (1992). Articulation disorders among speakers of Mandarin Chinese. *American Journal of Speech-Language Pathology*, *1*, 15–16. doi:10.1044/10580360.0104.15

Fletcher, S. G., McCutcheon, M. J., & Wolf, M. B. (1975). Dynamic palatometry. *Journal of Speech and Hearing Research*, *18,* 812-819. doi:10.1044/jshr.1804.812

Forrest, K. M., Weismer, G. G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84,* 115-123. doi:10.1121/1.396977

Gilabert, R., Barón, J., & Llanes, Á. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*, 367-395. doi:10.1515/iral.2009.016

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*, 338-365. doi:10.1093/applin/22.3.338

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36,* 664-679. doi:10.1016/j.wocn.2008.04.002

He, Y. (2014). Production of English syllable final /l/ by Mandarin Chinese speakers. *Journal of Language Teaching and Research, 5*, 742-750. doi:10.4304/jltr.5.4.742-750

Jongman, A., Waylad, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*, 1252-1263. doi:10.1121/1.1288413

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*, *33*, 103-115. doi:10.1044/jshr.3301.103

Lee, S. A. S., & Iverson, G. K. (2012). Stop consonant productions of Korean-English bilingual children. *Bilingualism: Language and Cognition*, *15*, 275-287. doi:10.1017/S1366728911000083

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *Journal of the Acoustical Society of America*, *130*, 4053-4062. doi:10.1121/1.3651816

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*, 124-139. doi:10.1044/1058-0360(2011/11-0009)

Nasukawa, K. (2004). Word-final consonants: Arguments against a coda analysis. *Proceedings of the Tohoku English Literary Society,* 47-53.

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *Journal of the Acoustical Society of America*, *118*, 2570-2578. doi:10.1121/1.2010407

Nissen, S. L., Hartshorn, J., Chase, B., Healy, H., Li, R., & Tanner, M. (2016). Acoustic and electropalatographic recordings from L2 learners of English [Audio recordings and binary data]. Provo, UT: Brigham Young University.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, *97*, 520-530. doi:10.1121/1.412278

Park, H., & de Jong, K. J. (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, *36*, 704-723. doi:10.1016/j.wocn.2008.06.002

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V.,…Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841-1852. doi:10.1093/brain/121.10.1841

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93-120. doi:10.1111/1467-9922.00071

Slobin, D. I. (1992). *The crosslinguistic study of language acquisition*. Hillsdale, NJ: Erlbaum.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *58*, 439-473. doi:10.1111/j.1467-9922.2008.00446.x

Yohiyuki, H., & Brandenburger, P. (1986). Predicting children's oral reading proficiency from selected acoustic and perceptual measures. *Journal of Communication Disorders*, *19*, 299-309. doi:10.1016/0021-9924(86)90036-5

APPENDIX A: ANNOTATED BIBLIOGRAPHY

Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition*, *13*, 263-278. doi:10.1017/S136672890999006X

Objective: This study examined the production of word-final consonant voicing by adult Koreans learning English and whether it is affected by age of acquisition and English experience. It also addressed the effects of age of acquisition and English experience on vowel duration and closure duration cues during voicing. Method: English-learning Koreans with varying ages of acquisition and English experience were put into three groups based on fluency level. Participants named pictures that elicited English vowels (/i/, /ɪ/, /æ/, /e/, /u/, /ʊ/) and word-final stops. Results: Age of acquisition and one's level of English experience affected word-final stop production in English. Closure duration was harder for L2 English learners to acquire than vowel duration, and it appeared to predict voiced word-final consonants. Participants who had started learning English 10 years previous still experienced difficulty producing word-final stop closure duration. Age of acquisition appeared to effect perceptual rating more than years of L2 experience. Conclusions: Age of acquisition may be a better predictor of accuracy of word-final consonant voicing in L2 learners than English experience. Vowel duration, closure duration, and release burst of a word-final stop are all important to English learners when judging word-final stops perceptually.

Boersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *The Journal of the Acoustical Society of America*, *127*, 1636-1644. doi:10.1121/1.3292996

Objective: To determine whether any differences exist in the perception of vowel duration of /v/ and /f/ between English and Dutch listeners. Also, to investigate whether vowel duration helps listeners discern /s/ from /z/.

Experiment 1

Method: The quality of a series of word-final fricatives (/z/, /s/, /f/, and /v/), were rated by 16 Dutch listeners and by 16 English listeners. Results: All listeners rated longer vowels

higher than short vowels in terms of fricative quality when the target was voiced. When the target was voiceless, shorter vowels were rated higher. English listeners' ratings were more affected by vowel length.

Experiment 2

Method: A new group of 16 Dutch listeners and 16 English listeners were asked to distinguish between productions of /v/ and /f/, and /s/ and /z/. Results: Both sets of listeners depended on both fricative voicing and vowel duration to distinguish between sounds. English listeners responded more with /v/ than Dutch listeners on the /v-f/ contrast portion and they relied more on vowel duration. In the /z-s/ contrast portion, there were no differences in the impact of vowel duration.

Conclusions: Dutch listeners and English listeners differ in how they utilize vowel duration when distinguishing between word-final fricative productions of the /v-f/ contrast, and the /z-s/ contrast. Vowel duration also effects the ability to distinguish between /s/ and /z/.

Fletcher, S. G., McCutcheon, M. J., & Wolf, M. B. (1975). Dynamic palatometry. *Journal of Speech and Hearing Research*, *18*, 812-819. doi:10.1044/jshr.1804.812

Objective: This article is an explanation of the function and instrumentation of the palatomer. The purpose of the palatometer is to measure contact between the tongue and the palate. A pseudopalate is molded from the patient's mouth and electrode contacts are installed inside the two plastic pieces that make up the pseudopalate. The patient's tongue completes an electrical circuit upon contacting the psuedopalate. A system of LEDs light up on a computer system when data is acquired so that contact patterns can be viewed immediately or at later time. The resulting data can also show the frequency of contact and average tongue contact with the pseudopalate across several repetitions. Conclusion: Repetitions of sounds vary in the way they are articulated in the mouth, and the palatometer can be useful when needing to observe speech articulation through linguapalatal contact.

Forrest, K. M., Weismer, G. G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*, 115-123. doi:10.1121/1.396977

Objective: Find a way to statistically organize word-initial voiceless obstruents. Method: Ten participants were recorded repeating the words paid, keen, tea, she, fought, pop, cot, tot, see, thought, pay, key, two, and fat in the carrier phrase "I can say _____, again". The voiceless obstruents were analyzed with Fast Fourier transforms and the four spectral moments of mean, variance, skewness, and kurtosis were calculated and plotted on a linear scale and a Bark scale. Results: The variable of time took the place of variance, due to the variance being unhelpful in distinguishing between obstruents. Mean was helpful in discriminating between /p/ from /t/, while skewness was not. Velar stops were easily identified using kurtosis measures. Using time measures, the linear scale was more successful when classifying voiceless stops than the Bark scale, but the Bark scale more accurately discerned between sibilants. Linear FFT calculations of mean, skewness, and kurtosis, successfully discriminated voiceless stops when the first 40 ms were analyzed. Neither scale accurately classified fricatives. Conclusions: Using spectral moments of kurtosis, mean, time, and skewness are helpful towards classifying word-initial voiceless obstruents.

Gilabert, R., Barón, J., & Llanes, Á. (2009) Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*, 367-395. doi:10.1515/iral.2009.016

Objective: To observe the effect of task complexity on different oral task performance within conversational interactions. Method: Sixty Spanish-speaking volunteers between the ages of 18 and 40 who studied English participated in three types of interactive tasks: decision-making, instruction-giving, and narrative. Within each task type were simple and complex levels of the task, all of which were done with a partner. The factor that changed within each learner's experience was task complexity. Each learner rated the task on a scale of task difficulty, personal stress, confidence, interest, and motivation. The number of confirmation checks and comprehension checks were also noted. Results: The variables rated by learners that were effected by task type were difficulty, stress, and

confidence. No effect was noted on interest and motivation. The three more complex versions of each task were rated to be more difficult and the complex narrative task was rated as the most difficult task, as well as the task that most affected confidence levels. Repairs that were made during the tasks were more numerous in the complex tasks. Conclusions: Changing task complexity can potentially increase interaction in an interactive task. As a task becomes more complex, language learners are more precise, they reorganize the delivery of their message, and they work towards being successful in the task. Increasing the complexity of a task may improve interaction, which can assist language learners.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*, 338-365. doi:10.1093/applin/22.3.338

Objective: The aim of this study was to investigate the effect of one's native language of Mandarin Chinese on his ability to learn English syllable-final codas. Method: Spontaneous speech was elicited from three speakers of Mandarin Chinese who each had two separate one-hour-long recorded interviews in English that were spaced six months apart. These interviews were transcribed phonetically and the syllable-final codas were analyzed. Results: Changing place or manner of articulation was the most common error in single codas. In two-member codas, inserting vowels was the most common error. The most common error in three-member codas was the deletion of one or more consonants. Conclusions: Depending on coda length, native Mandarin speakers make changes to syllable-final codas in the English language, whether it is changing place or manner of articulation, inserting vowels erroneously, or deleting consonants.

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*, 664-679. doi:10.1016/j.wocn.2008.04.002

Objective: To observe native English and Mandarin listeners' word identification patterns in both voiced and voiceless word-final stop consonants in the English language. Method:

Native English (NE) speakers and native Mandarin (NM) speakers were recorded producing words including minimal pairs like 'cab', 'cap', 'face', 'fez', 'peas', 'peace', 'take', and 'tech', which were then listened to and the words were identified by another group of NE and NM speakers. Another group of NM and NE speakers rated the first group's accentedness. Results: NM listeners were better than NE listeners at distinguishing words produced by NM talkers. NM listeners understood NE speech better than they understood NM speech. Conclusions: Nonnative speech may be comprehended better by nonnative listeners than by native listeners. The acoustic cues utilized by NE and NM listeners may differ when identifying contrasts in voicing produced by low-proficiency NM speakers.

Jongman, A., Waylad, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*, 1252-1263. doi:10.1121/1.1288413

Objective: This study sought to determine whether the place and manner of articulation of fricatives could be identified based on spectral properties, amplitude, and noise duration. Method: Twenty adult participants were recorded saying carrier phrases that included fricatives at the beginning of each target word. Spectral peak location, spectral moments, locus equations, root-mean-square amplitude, and relative amplitude were all measured. Results: Spectral peak location and spectral moments successfully identified place of articulation in fricatives. Observing amplitude also helped to identify different places of articulation in fricatives. Conclusions: Place of articulation in fricatives can be distinguished by spectral peak location, spectral moments, normalized amplitude, and relative amplitude.

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*, *33*, 103-115. doi:10.1044/jshr.3301.103

Objective: This research evaluated whether listener experience due to clinical training influences perception of voice quality. Method: The voices of 18 males with voice disorders and 18 males with normal voices were randomly presented twice to each of five naïve listeners and five expert listeners. All participants rated voice pair samples based on

their similarity to one another and then listened to and rated them again in terms of fundamental frequency, breathiness, roughness, abnormality, and jitter. Results: When rating the voices, the expert listeners relied most heavily on measurements of breathiness and roughness, while naïve listeners relied more on fundamental frequency. Conclusions: Naïve listeners depend on fewer variables to make perceptual judgments on vocal quality. Also, the more experience that a listener has, the more his ratings will differ from that of other expert listeners. Naïve listeners agree more in terms of what dimensions are most important when judging perceptual quality.

Lee, S. A. S., & Iverson, G. K. (2012). Stop consonant productions of Korean-English bilingual children. *Bilingualism: Language and Cognition*, *15*, 275-287. doi:10.1017/S1366728911000083

Objective: This study evaluated at what time and through what mechanisms Korean-English bilingual children create complete separate phonetic systems for Korean and English, by analyzing differences in voice onset time and vowel-onset fundamental frequency in stops. Method: A group of children ages 10, 15, and 5, some bilingual in Korean-English, others monolingual in English or Korean, were recorded producing target words in a carrier phrase. Results: Monolingual Korean children produced a higher fundamental frequency at vowel-onset than bilingual children. The bilingual children produced a longer voice onset time for stops in Korean than monolingual children. Conclusions: Five-year-old Korean-English speaking children utilize assimilation in their phonetic system, while ten-year-olds utilize dissimilation and assimilation.

Levi, S. V., Winters, S.J., & Pisoni. D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America*, *130*, 4053-4062. doi:10.1121/1.3651816

Objective: To observe how listeners are able to identify familiar speakers depending on the context of language. Method: Native German speakers who were L2 learners of English were recorded saying words in both German and English. Native English speakers were trained on identifying the speakers' voices and were then split into two groups, one where they listened to German and the other where they listened to English.

All the listeners were then asked to identify the speakers but of recordings of English words spoken by both familiar and unfamiliar speakers. Results: The participants who were trained on English were more accurate when identifying speakers than those who were trained on German. For the German-trained listeners, familiar talkers were not more intelligible in English than unfamiliar talkers. Conclusions: Listeners are not helped in identifying talkers in word recognition tasks when the language differs from the trained language.

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*, 124-139. doi:10.1044/1058-0360(2011/11-0009)

Objective: This study analyzed the differences between perceptual ratings made on a visual-analog scale of children's speech production by speech-language pathologists and those rated by people with no clinical experience. Method: Forty-two listeners were split into two groups: inexperienced listeners and experienced listeners. The inexperienced listeners ranged from age 18 to 50 years, while the experienced listeners ranged from age 26 to 59 years. The experienced group had on average, 13 years of listening experience. All participants listened to young children's (ages two through five) phoneme productions and rated the quality of each production on a visual-analog scale. The listeners had to distinguish between /d/-/g/, /s/-/θ/, and /t/-/k/. Results: The experienced listeners usually exhibited two or modes in the way they gave ratings, while inexperienced listeners' modes were not as numerous nor as strong in the distribution of their ratings as the experienced listeners. The ratings of the experienced raters were also more reliable than the inexperienced listeners in every stimulus set. The intra-rater reliability for the experienced listeners was higher than for the inexperienced listeners. Conclusions: Listening experience that one gains in a clinic aids in becoming more consistent than inexperienced listeners when judging speech samples. When distinguishing between phonemes, experienced listeners are more likely to judge a sound as sounding like /θ/, /k/, and /g/. The inexperienced listeners were more likely to rate sounds as phonemes that occur more frequently in language. The changes in properties of

sound productions influenced experienced listeners' ratings more than inexperienced listeners' ratings. Overall, it was found that different levels of listening expertise influence perceptual speech ratings.

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America,* 118, 2570-2578. doi:10.1121/1.2010407

Objective: Analyze the production and differences of voiceless fricatives in children and adults by observing spectral moments and acoustic characteristics. Method: Native English speaking children were recorded as they produced words with fricatives in the initial position in the carrier phrase "This is a _____." Results: As age increased from three to five years old, classification rates improved. For spectral kurtosis, slope, mean, variance, and skewness, place of articulation was statistically significant. Conclusions: Spectral variance successfully discerns place of articulation in fricatives, while spectral measures of slope, mean, and skewness distinguish non-sibilant from sibilant fricatives. Children begin to exhibit more advanced contrast at four years of age as skewness is observed, and they demonstrate improvement in classification rate beginning at five years old. There is a "fine-tuning" of children's articulation of sibilant fricatives that takes place as they mature.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, 97(1), 520-530. doi:10.1121/1.412278

Objective: The articulation and development of fricatives in children's speech compared to adults' speech is examined in this study. Another objective is to determine what effects age has on spectral moments within different vowel contexts. Method: Adult and child participants were recorded producing /s, ʃ, t,/ and /k/ combined with different vowels. Spectral moments were then computed and analyzed. Results: First moments in stop bursts did not differ among ages, while first moments for fricatives did differ across consonant effects in adults. Conclusions: There are some speech sounds that develop more quickly than others. Spectral moments between /s/ and /ʃ/ differ more in adult

speech than in children's speech. Adults are more skilled at differentiating mouth movements than children.

Park, H., & de Jong, K. J. (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, *36*, 704-723. doi:10.1016/j.wocn.2008.06.002

Objective: The aim of this study was to examine the identification of English anterior obstruents by Korean L2 learners of English, in order to determine how much L1 categories are used in L2 identification. Method: Forty Korean participants with seven or more years of English experience listened to nonsense words that contained English obstruents /b, p, d, t, v, f, ð/ and /θ/. Each subject labeled each stimulus with Korean orthography, and Roman and IPA symbols. They also rated to what extent their label sounded like the stimulus on a Likert scale of one to seven. Results: Most of the English obstruent stimuli were labeled consistently with similar Korean characters as follows: /p/ was consistently labeled with Korean /pʰ/, /d/ and /ð/ with /t/, /t/ with /tʰ/, /v/ with /t/ and /p/, /f/ with /pʰ/ and /p'/, and /b/ was often labeled with Korean /p'/ and /p/. The fricative /θ/ was not consistently labeled with one Korean sound. The mapping of English voiceless stops and Korean aspirated stops were directly correlated, while the voiced stops and fricatives of English were less directly correlated with Korean phonemes. The Roman labeling task proved to be difficult and the results were not consistent. Conclusions: Some speakers develop L2 categories by using their familiar L1 categories, while others need to create a new linguistic category.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., …Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841-1852. doi:10.1093/brain/121.10.1841

Objective: These two studies address whether the age of acquisition influences L2 speakers' proficiency level.

Experiment 1

Method: Nine Italian men who were moderately proficient in English had their regional cerebral blood flow measures taken while listening to a story in their native language, in English, and an unfamiliar language played backwards. They were then asked listening comprehension questions. Results: high-proficiency speakers of English with late acquisition had similar brainwave activation patterns when listening to English, Italian, and backward language. The low proficiency speakers of English with late acquisition did not show brain activation in as many parts of the brain as the other high-proficiency speakers of English with late acquisition.

Experiment 2

Method: Twelve fluent Spanish-Catalan speakers had their regional cerebral blood flow measured as they listened to stories in Spanish, Catalan, and those languages played in reverse. The brain pattern results for the high-proficiency early acquisition participants were similar to those from Experiment 1. Also, some regions of the brain activated in Catalan but not in Spanish, and vice versa.

Conclusions: In low proficiency late acquisition individuals listening to stories in L1 and L2, different brain activity occurs. In high proficiency listeners with different ages of acquisition, there were not any major differences seen between brain activity when listening to stories in L1 and L2. Proficiency level causes differences in brain activity more than age of acquisition does when listening to stories in L1 and L2. Also, the difference between two languages does not influence overlap when an individual is proficient in both languages.

Skehan, P., Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93-120. doi:10.1111/1467-9922.00071

Objective: To study the effects that amount of structure to a task and processing have on fluency, complexity, and accuracy while performing a narrative task. Method: Forty-seven participants who spoke various first languages and were studying English as a second language at a university in London were assigned to one of two videos under one

out of four conditions: (1) narrating a video while they watched it (the most demanding), (2) hearing an outline of the video and then narrating it as they watched it, (3) watching the video first and then narrating it as they watched it a second time, and (4) watching the video and afterward giving a narrative of the story (the least demanding). Each narration was coded for complexity, accuracy, and fluency. Results: Between the two videos, the complexity and accuracy did not differ significantly, while the fluency measures were higher for one video than the other. The condition when they watched the video and then retold what happened proved to elicit the most complex and accurate language from speakers. The four fluency measures correlated strongly with one another. Non-simultaneous tasks proved to help speakers produce more complex language. However, being given the storyline before watching the video had no effect on language complexity. Conclusions: Sequential and clearly structured tasks lead to more fluent language production than less structured tasks. Accuracy of performance is only significantly affected by task structure when combined with pre-task preparation. More complex language is not affected by task structure, but it is affected by conditions of the task, including processing demands.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *58*, 439-473. doi:10.1111/j.1467-9922.2008.00446.x

Objective: To determine how different aspects L2 learners' language are influenced by task structure. Method: One-hundred L2 intermediate learners of English with a variety of native languages were each asked to give English narratives of two cartoons. Some of the cartoons allowed loose narrative structure and others did not. Some also allowed for background events and others did not. Results: Participants that described the cartoons with both background and foreground information exhibited a higher level of language. When given a tight structure versus a loose structure story, language performance was significantly higher on tight structure tasks. Conclusions: The complexity of narrative tasks affects L2 learners' oral language performance. Tighter structure directly correlates with language performance.

Tsao, F., Liu, H., & Kuhl, P. K. (2006). Perception of native and nonnative affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*, 2285-2294. doi:10.1121/1.2338290

Experiment 1

Objective: Observe how Mandarin and English adult speakers perceive Mandarin affricate-fricative consonants. Method: Eighteen English speakers and 18 Mandarin speakers were required to discriminate between phonemes and tell whether pairs of phonemes were the same or different. Results: Mandarin speakers utilized amplitude rise time and frication duration to distinguish contrasts between phonemes, and they scored 10% higher than English speakers, with 77.62% accuracy. Conclusions: It is easier to perceive affricate-fricative consonant contrasts in a native language than in a nonnative language.

Experiment 2

Objective: To analyze patterns of change in phonetic perception in both English and Mandarin in infants. Method: Sixty-nine American and Taiwanese infants, between 6-8 months or 10-12 months, were observed as they responded to Mandarin affricate-fricative consonants. After each sound was played, it cued a visual reinforcer on either side of the infant's head. The infant's attention was brought back to mid-line by an assistant using toys. Results: English-learning infants' fricative-affricate consonant discrimination declined with age and Mandarin-learning infants improved with age. Conclusion: Within the first year of life, native language learners can more easily discern differences between native phonemes than nonnative listeners.

Experiment 3

Objective: Observe patterns of change in native perception of affricate-fricative consonants in infants. Method: American infants 6-8 months old and 10-12 months old were observed as they discerned between the phonemes /ʃi/ and /tʃi/ using the conditioned head-turn technique. Results: The 10-12-month-old infants performed better than the 6-8 month infants. When compared to Experiment 2, the older English-learning infants could detect differences more accurately in native sounds than in nonnative sound, and the

younger infants were no better at one language than the other. Also, Mandarin and English-learning 11-month-old infants were equally skilled at discriminating native sounds. Conclusion: Among infants from different language backgrounds, the rate of perception of native affricate-fricative contrasts is parallel.

Yohiyuki, H., & Brandenburger, P. (1986). Predicting children's oral reading proficiency from selected acoustic and perceptual measures. *Journal of Communication Disorders*, *19,* 299-309. doi:10.1016/0021-9924(86)90036-5

Objective: Research was conducted to determine whether perceptual analysis and reading error analysis help in predicting children's reading proficiency and whether gender affects children's reading proficiency. Method: Eighteen male and 18 female fifth and sixth-grade elementary school students were recorded reading two passages twice. Both expert and naïve listeners rated their reading proficiency. The author analyzed reading errors that the participants made in six categories: substitutions, omissions, insertions, reversal, repetitions, and grammatical conversations. Results: Predictions of oral reading proficiency were fairly accurate. Ratings decreased as more errors were made. The most common reading errors made were substitutions (34.7% of all errors), repetitions (24.8%), and insertions (17.2%). The girls performed better than the boys on the oral reading proficiency portion. Conclusions: Predicting reading proficiency through perceptual ratings is effective, especially when reading pace and reading errors are considered. The reading errors that most commonly occur between the ages of 10-12 are substitution and repetition errors. Oral reading proficiency is higher on average in girls than in boys between the ages of 10-12.

APPENDIX B: CONSENT FORM

Consent to be a Research Participant (Adult listener)

Introduction
The purpose of this study is to collect listener evaluations of the speech clarity of three production tasks for second language learners of English. This experiment is being conducted under the supervision of Shawn Nissen, Ph.D., an associate professor in the Department of Communication Disorders at Brigham Young University. You have been invited to participate because you are an adult English speaker with no known history of a hearing impairment.

Procedures
Participation in this study will involve one visit of approximately sixty minutes, which will take place in a research laboratory in the John Taylor Building at BYU. You will be asked to listen to everyday English words and sentences spoken by adults and respond regarding your perception of the pronunciation clarity of their speech.

Risks/Discomforts
There are minimal risks for participation in this study.

Benefits
There are no direct benefits to participants. However, it is hoped this study will provide understanding that may assist clinicians and researchers in developing more effective approaches to second language learning.

Confidentiality
All information provided will remain confidential and will be reported only as group data with no identifying information. All data, including records of your listening responses, will be kept on password-protected computers in a locked laboratory and only those directly involved with the research will have access to them.

Compensation
You will be compensated $10 per hour for your participation in this study.

Participation
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate without penalty.

Questions about the Research
If you have questions regarding this study, you may contact Shawn Nissen, Ph.D., at (801) 422-5056 or shawn_nissen@byu.edu.

Questions about your Rights as Research Participants

If you have questions regarding your rights as a research participant, you may contact the BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT, 84602 or at (801) 422-1461.

I have read and fully understand the consent form. Any questions have been answered to my satisfaction. I give my consent to participate in this research.

Signature:_____   Date:_____

Printed Name: _____

APPENDIX C: TASK TYPE PROMPTS

Paragraph Readings

Last spring Tom and Kate went to the beach with their dog. After removing his leash, they watched him climb a steep mound of dirt and leap in the air. Tom split a piece of beef and threw it for the dog to catch in his teeth. They found a good seat in the sand and began to read a letter from a dear friend. After taking a peek at one page, Tom's head began to tilt back and he began to sleep. Kate started to think, I wonder if Tom covered his skin to screen out the sun. She decided to speak, "Please wake up". It was too late. Tom had deep burns on his cheek, the skin had started to peel. They decided to clean the sand from their feet and walk to the thrift shop at the top of the street. The shop not only sold cream to treat his burns, but also had kilts and shirts made of fleece. Tom began to preach, "Those poor sheep, I hope they don't freeze."

One spring evening, my mother made a fruitcake to sell at the thrift shop at the end of the street. First, she split a peach in half and began to peel away the skin. She smiled with her teeth as she began to read the instructions. Next, she told me to clean the fruit to remove any dirt. With the bowl at a tilt she added cream to the treat. My dear father rose to his feet and kissed her on the cheek. She then gave him a leftover piece of beef to eat. He said, "I need to feed those sheep before I go to sleep." He quickly rose from his seat, took a leap across the room, and rushed through the screen door. Mother said "I think we should put him on a leash." With the cake in the oven, she began to speak and preach about her memories swimming in the deep ocean far from the beach and climbing a steep mountain peak to watch men freeze while dancing in fleece kilts.

APPENDIX C: TASK TYPE PROMPTS

List of Fairytales

1. 掩耳盗铃 Plugging One's Ears While Stealing a Bell

2. 塞翁失马 Blessing in disguise

3. 狐假虎威 The fox assuming the majesty of the tiger

4. 狼来了 The wolves are coming

5. 牛郎织女 Chinese Romeo and Juliet

Open-Ended Questions

1. Describe for me what happened on the day that you arrived in the United States.

    a. Tell me more about what happened when…

2. Share with me why learning English is so important to you.

    a. Tell me more about how English will create opportunities for you.

3. Tell me about a favorite memory when you were young.

4. Tell me about your hometown.

5. Tell me about a recent vacation.

6. Tell me about how living in the United States is different than your home country.

7. Tell me a story about a time you had fun with your family or friends.

8. Tell me what you enjoy doing in your free time. Why?

9. Tell me about your education before coming to the ELC.

10. Tell me about your favorite subject in school and why.