Age Effect on the Acquisition of Second Language Prosody

Becky H. Huang & Sun-Ah Jun
University of California, Los Angeles

ABSTRACT

This study explores the effect of age on the acquisition of second language prosody by comparing three groups of native Mandarin speaking immigrants (N=10 each), varying in age of arrival, to native American English controls (N=10). Speech samples of participants reading a paragraph were subjected to three prosodic analyses: speech rate, native speakers' judgment of foreign prosody, and analyses of intonation and prosodic groupings. Preliminary findings show that second language prosody is significantly influenced by the age of arrival factor, but the strength of the age constraints varies among different aspects of prosody.

BACKGROUND INFORMATION

The age-related decline in ultimate second language (L2) attainment is perhaps one of the most controversial topics in the L2 acquisition research due to its theoretical and practical implications. Among the various linguistic features, the acquisition of L2 phonology is potentially the least controversial for an age of learning effect. In fact, Scovel (1988) argued that the age effect exists only for phonology since the ability to master the sound patterns of a second language is susceptible to neurological development. Although only a few empirical studies have examined the age effect across linguistic domains, the results seem to suggest that immigrants' ages of arrival (AoA) in the L2-speaking country, generally used as the proxy variable for the age effect, constrain the phonology domain more than the morphosyntax domain (Flege, Yeni-Komshian, & Liu, 1999; Huang, 2008; Oyama, 1973).

Extant studies on the ultimate proficiency in the production of L2 sounds have predominantly focused on segmental properties. These studies reveal a firm relationship between AoA and L2 pronunciation accuracy. For example, Flege and colleagues (Flege et al, 1999; Flege, Munro, & MacKay, 1995) found Italian native speakers who arrived in the United States at a younger age produced English (L2) vowels and consonants more accurately than those who arrived later. However, relatively little research has focused on L2 acquisition of suprasegmental properties despite their crucial role in language development, comprehension, and production (Chun, 2002).

Suprasegmental or prosodic features[1] have been argued to differ fundamentally from segmentals (Lehiste, 1970). The importance of prosody in language learning and processing has received empirical support in several research areas. Specifically, in the first language acquisition literature, studies have shown that children acquire intonation earlier than consonants and vowels (Golinkoff, 1983; Li & Thompson, 1977). Prosodic cues have also been found to play an important role in facilitating infants' word learning (Shi & Moisan, 2008), assisting adults in solving ambiguous syntactic structures (Kielgaard & Speer 1999; Millotte, Wales, & Christophe, 2007; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991), and helping speakers achieve discourse coherence via the use of increased pitch range to signal topic shift (Hirschberg & Pierrehumbert 1986, Wennerstrom, 1998). Additionally, prosody has been shown to contribute to listeners' pronunciation judgment (Magen, 1998), and in a few studies, even found to weigh more than the segmental properties (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992).

Despite the importance of prosody, relatively little research has been devoted to the acquisition of L2 prosody, and even scarcer is research on the effect of age on the production of L2 prosody. In a recent study, Trofimovich & Baker (2006) examined the effect of L2 experience on the acquisition of prosody by recording Korean speakers' production of English (L2). The recordings were subjected to acoustic measurements (i.e., speech rate, pause duration, pause frequency, stress timing, and tonal peak alignment) and the degree of foreign accent was rated by native English speakers based on low-pass filtered speech recordings. They found that participants' L2 experience, as measured by participants' length of stay in the US, was correlated with stress timing analysis and foreign accent rating, with more experience resulting in more native-like production. However, L2 experience was not correlated with the acquisition of peak alignment and a higher speech rate. Neither did L2 experience help reduce pause frequency and duration. However, this study only included post-pubescent L2 learners. It thus did not address the experience effect in learners with different ages of immersion, an issue addressed by our current study.

Narrowing down the L2 prosody production literature to studies on age effect, the identified short-term laboratory learning/training research suffers from methodological shortcomings and the majority of the long-term immersion learning studies are found to center upon the global foreign accent (Asher & Garcia, 1969; Flege, MacKay, & Meador, 1999; Flege et al, 1999; Neufeld, 1978; Oyama, 1976). Specifically, in a short-term laboratory study, Neufeld (1978) taught Anglophone adult participants to produce

[1] In this paper, "prosody" is defined as the prominence relations and grouping of subunits in an utterance, such as stress, rhythm, and intonation. Acoustically, prosody is realized through variation in duration, loudness, and pitch.

lengthy sequences in three languages, one of which was the tone language Mandarin Chinese. The participants were able to achieve native or native-like pronunciation of the Mandarin tones, suggesting that their perceptual and motor systems were still malleable despite their late start. However, participants were instructed to reproduce the sequences without being provided the rules or meanings of the sound sequences. Without such knowledge, the participants' success could have been attributed to their vocal imitative skills, and not their ability to master the prosody of a second language.

In contrast, long term immersion studies, such as those conducted by Flege and colleagues with immigrants from various first language (L1) backgrounds (e.g. Asher & Garcia, 1969; Flege, MacKay & Meador, 1999; Flege et al., 1999; Oyama, 1976), have been methodologically more rigorous than the short-term studies. This line of research confirms the age effect found in segmental pronunciation research by showing that late arrivals were perceived to have a stronger global foreign accent than early arrivals based on their production of L2 sentences. However, global foreign accent is not a "clean" measure of prosody, but a mixed product of segmental and suprasegmental deviations (Magen, 1998) as well as perhaps other non-linguistic variables such as confidence level (Piller, 2002). To obtain an accurate measure of prosody production without the interference of segmental information, low-pass filtering technique has thus been used by some researchers in the investigations of L2 prosody (e.g., Munro, 1995; Trofimovich & Baker, 2006). This promising technique, however, has not been utilized in studying specifically the age effect on L2 prosody acquisition.

Finally, a remaining issue in the research on the age effect on L2 prosody, and also in the L2 prosody research in general, concerns the methodology for characterizing prosodic patterns. Research on segmental properties has well-established quantitative and qualitative analysis methods that are not only relatively straightforward but readily available. In contrast, research on prosodic properties has been limited to quantitative measurements of pitch and duration on a certain position in an utterance, and methods of analyzing prosody qualitatively, though well developed in theoretical linguistics (Beckman & Hirschberg 1994, Beckman & Ayers-Elam 1994), have only been applied to a few L2 prosody empirical studies (Jun & Oh, 2000; Ueyama & Jun, 1998; Wennerstrom, 1998). The current study thus addresses this issue by applying a qualitative analysis method in the investigation of the age effect on L2 prosody. The development of a quantitative analysis method of prosody is in progress and will not be included in the current paper.

STUDY GOALS & RESEARCH QUESTIONS

To summarize, the literature review above reveals the importance of researching the age effect on L2 phonology, the dearth of studies specifically focusing on L2 prosody production, and the methodological issues of L2 prosody research. The current study thus aimed to fill the gap and create a fuller understanding of the age effect on L2 phonology by providing an empirical exploratory report of the age effect on L2 prosody. The current study also utilized inventive analytical approaches, specifically the low-pass filtering technique and the phonological model of prosodic transcription (the Mainstream American English Tones and Break Indices (MAE_ToBI)), to examine the influence of age of arrival variable on Mandarin-speaking immigrants' acquisition of various aspects of L2 prosody. The findings will not only shed light on theories of second language acquisition, but also have direct implications for language education. Specifically, the study will inform us of the specific prosodic features that pose challenges to ESL learners who have started their immersion experience at varying ages. Educators can take advantage of the findings and accordingly tailor the curriculum and instruction for ESL learners of different ages. The findings will also be useful in policy-making decisions regarding the optimal age for teaching a foreign/second language.

The overarching research questions pursued by the current study are as follows: ***Is there an age effect on the ultimate outcome of various aspects of second language prosody? If so, which prosodic features are impacted by the age effect?*** To answer the research questions, the current study included Mandarin-speaking immigrants with differing ages of arrival in the United States as well as a group of native speaker controls. Three approaches were employed to analyze different aspects of prosody, including 1) analysis of speech rates; 2) comparing native speakers' rating of low-pass filtered speeches, and 3) labeling and examining the frequency and locations of intonational categories and prosodic organizations.

METHOD

Participants

Selected participants included three groups of Mandarin-speaking English learners who varied in their Age of Arrival (AoA, Range=5-27). All participants spoke Mandarin as their first language, have lived in the United States for at least five years, and had received either none or only English instruction in a regular foreign language classroom prior to their arrival. All participants also held a college degree or were current college students, and have never been diagnosed with hearing problems, language disorders, or learning disabilities. Based on their AoA, the three learner groups were labeled as "Child

Arrival" (AoA=5-9 years old), "Adolescent Arrival" (AoA=12-17 years old), and "Adult Arrival" (AoA=20-26 years old). All three learner groups had lived in the U.S. for approximately ten years (range = 8-18), and the length of residence (LoR) did not differ among those groups, $F(2, 21)=1.15$, $p=0.34$. However, participants' current ages (Age) were significantly different among AoA groups, $F(3, 30)=16.16$, $p <.000$. Given the linear dependency of AoA, LoR, and Age variables, it was not possible to control for both length of residence and current age.

A group of ten native speakers (NS) of American English also participated and served as the control group. All NS participants spoke English as their first language, and had had only foreign language exposure in their high school classroom. They were affiliated with a university as students or staff. See Table 1 for a summary of the participants' demographic information.

Table 1. Demographic Information (Age, AoA, LoR) by Group

	Native Speaker (N=10)	*Child (N=10)*	*Adolescent (N=10)*	*Adult (N=10)*
AGE[a]	25.00 (5.81)	20.11 (2.03)	28.75 (5.65)	35.14 (2.19)
AoA[b]	NA	7.67 (5-9)	13.75 (12-17)	23.43 (20-26)
LoR[c]	NA	12.33 (2.74)	13.75 (3.24)	11.64 (2.14)
Gender	6f, 4m	7f, 3m	6f, 4m	6f, 4m

a.Current age (standard deviation in parenthesis); b. Age of Arrival (range in parenthesis); c. Length of residence (standard deviation in parenthesis).

Instrument

Previous research on L2 speech production revealed that task types played a role in production assessment. Specifically, words and sentences tended to overestimate participants' proficiency level compared to paragraphs and spontaneous speech (Moyer, 1999). To circumvent this pitfall, an elicitation paragraph from the Speech Accent Archive website (http://accent.gmu.edu/) was used for the current study.[2] The English paragraph consists of sixty-nine words, seventy-seven syllables, and 4 sentences (See Appendix I for the paragraph). The paragraph encompasses practically the full inventory of American English vowels and consonants. However, it is composed of all declarative sentences with straightforward imperative meanings as opposed to various sentence types and pragmatic and discourse meanings, and is thus quite limited in the realization of intonation patterns.

[2] We thank Steven Weinberger at George Mason University for giving us permission to use the paragraph.

Although participants' reading ability was suggested in a previous study to be a confounding factor in a read-aloud task (Munro & Derwing, 1995), it was unlikely the case for the current study because words in the paragraph are all simple common words. Participants were also given one minute to familiarize themselves with the script before they started recording.

Procedure

All Mandarin-speaking participants were recruited through a variety of venues, and were individually tested in a quiet room at the university or their private residences. Native speaker controls, on the other hand, were all students and staff from the same university and tested in a university laboratory room.

Each participant was given one minute to review the elicitation paragraph before the recording, and was instructed to read at their natural pace. They then read the paragraph twice into a high quality head-mounted microphone (Shure SM 10A) with the Audacity recording program. At the end of the language testing session, all participants filled out a survey about their demographic information. The survey data were not included in the current study.

DATA ANALYSIS & RESULTS

Speech Rate Analysis

Speech rate was calculated as the ratio derived by dividing the number of syllables in the paragraph by the total duration of the speech file, including pauses during the production of the paragraph. The ratio (number of syllables produced per second) was used as the dependent variable in the following ANOVA analysis.

A one-way ANOVA revealed a significant group effect, $F(3, 36)=6.14$, $p<.000$, and Tukey's post hoc tests showed that only the *Adult Arrivals,* not *Child* and *Adolescent Arrivals*, had a significantly slower speech rate than the *NS* ($p<.000$). Among the three learner groups, *Adult Arrivals'* speech rates were also significantly slower than *Child Arrivals* ($p<.000$), but there was no significant difference between either *Child* and *Adolescent* or *Adolescent* and *Adult Arrivals*. Current findings corroborated previous research on the age effect on speech rate by Guion, Flege, Liu & Yeni-Komshian (2000) and Mackay, Flege, & Imai (2006).

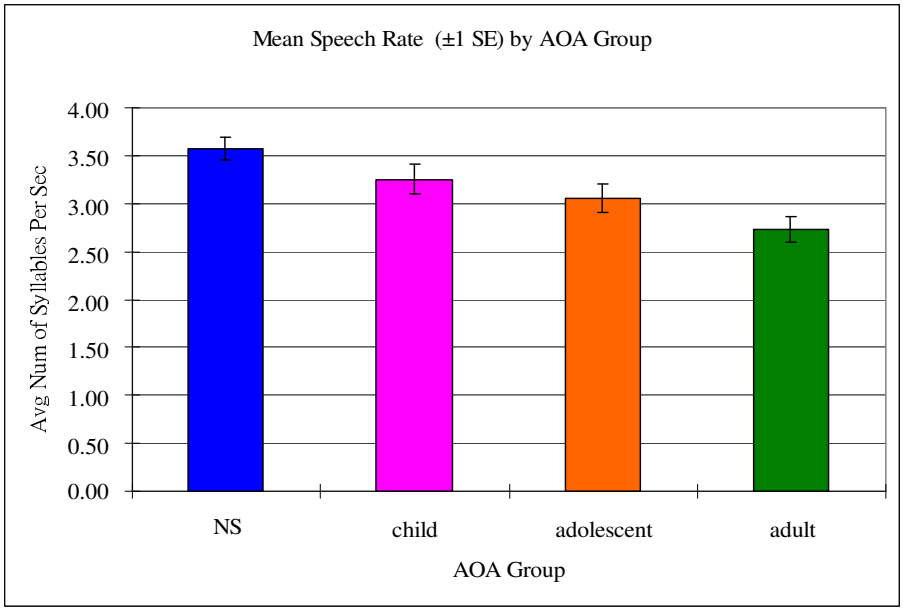

FIGURE 1. Group means for speech rate (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

Analysis of Foreign Prosody Rating via Low-Pass Filtered Speech

The recordings were first segmented to include one second of silence before and after the reading of the paragraph. The segmented paragraphs were then band-pass filtered (cutoff frequency=50Hz and 450Hz) to remove the segmental information while preserving the prosodic information. To preserve the amplitude of the original signal, especially the vowels, the intensity curve of the original file was calculated and the filtered files were multiplied by the intensity curve. The procedure was conducted with the PRAAT software (http://www.fon.hum.uva.nl/praat/). The forty filtered files were then divided into two blocks with equal numbers of files from each participant group. The twenty files within each block were fully randomized via the MATLAB program.

Twelve American English native speakers between the age of 18-33 (M = 24) were recruited from a university in Southern California to rate the prosody of the filtered files. All raters reported normal hearing and spoke standard American English as their first language. There were equal numbers of females and males. Half of the raters were linguistics majors or former ESL teachers, and half of them were naïve raters with no linguistics background.

The filtered files were played on a laptop (ThinkPad x60s) via headsets (Sennheiser Model HD212), and were presented at a comfortable listening level. The raters were given a written copy of the paragraph against which they could compare the filtered speech. The raters were told that they would listen to filtered speeches produced

by both native speakers and nonnative speakers, but the proportions of each population were unspecified. They were instructed to judge only the intonation of the speech and disregard segmental information, if they heard any, and encouraged to use the full scale for rating.

The raters listened to each filtered file once and assigned a score of 1-9 (1=strong foreign intonation; 9= native-like intonation) to indicate the degree of foreign intonation for each recording. The inter-stimuli interval is one second. Before the actual rating sessions, they were given five practice trials to familiarize themselves with the filtered speech and received feedback on their ratings of the practice trials. Each rating sessions lasted about ten minutes, and s short break was provided between the two sessions.

Results showed that the twelve raters' ratings were all highly and significantly correlated with one another. The Intraclass correlation obtained for the raters was also high (R=0.93, $p<.000$), justifying the use of an average rating as the dependent variable for each participant. Figure 2 presents the mean ratings of foreign prosody (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).[3] A one-way ANOVA revealed a significant group effect, $F(3, 36)=11.15$, $p<.000$. Tukey's tests comparing the learner groups with the *NS* further revealed that both *Adult* and *Adolescent Arrivals* received significantly lower ratings than the *NS* ($p<.000$ and $p<.05$, respectively). *Child Arrivals*, however, were not significantly different from the *NS*. Among the learner groups, *Adult Arrivals* were perceived to have a stronger foreign prosody than *Child Arrivals* ($p<.000$). The average ratings between *Child* and *Adolescent Arrivals* and between *Adolescent* and *Adult Arrivals* were not significantly different.

[3] The overall mean rating is lower (Range: 3.45 ~ 6.86) than that reported in the literature based on unfiltered speech because, as indicated by the low rating for the NS group (M=6.86), rating low-pass filtered speech is much harder than rating unfiltered speech,

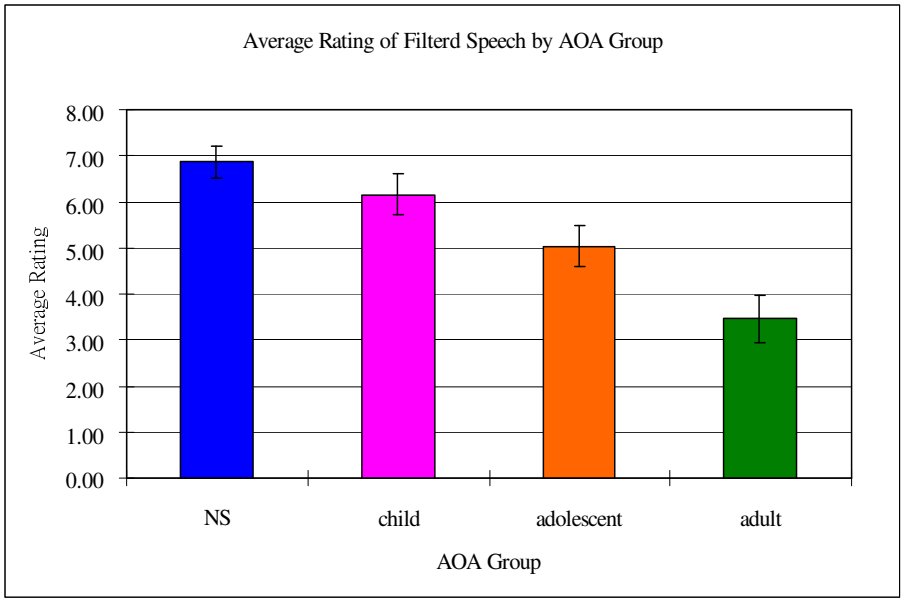

FIGURE 2. Group means for foreign prosody rating (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

TOBI Labeling & Analysis

The study adapted the Autosegmental-Metrical phonology model of English intonation (Bruce, 1977; Beckman & Pierrehumbert 1986; Ladd, 1996; Liberman, 1975; Pierrehumbert, 1980) as the theoretical framework and utilized the Mainstream American English Tones and Break Indices (MAE_ToBI) prosodic transcription conventions (Beckman & Hirschberg 1994, Beckman & Ayers-Elam 1994) to label and analyze the prosodic groupings and intonation patterns of the speech production. The current MAE_ToBI (1994) includes an inventory of nine pitch accent types (H*, !H*, H+!H*, !H+!H*, L*, L+H*, L*+H, L+!H*, L*+!H) to mark prominence, two phrase accent types (H- & L-) to mark small intermediate phrase boundaries, and two boundary tone types (L%, H%) to mark large intonation phrase boundaries. The ToBI conventions capture the categories of English intonation ('To' part) as well as the hierarchy of English prosodic organization ('BI' part). Specifically, pitch accents are associated with stressed syllables, phrase accents are properties of intermediate phrases, and the boundary tones are associated with intonational phrases. Although the MAE_ToBI conventions includes both tones and break indices conventions, the current study only labeled the tones, i.e., the pitch accents, phrase accents, and boundary tones.

For the current paper, all ten native speaker files and seven randomly selected speech files from each Mandarin-speaking learner group (N=21) were analyzed (N=31 total). Since each participant read the paragraph twice, the researchers first listened to

both readings and selected the one with higher fluency and better recording quality. In most cases, the second recordings were selected. Each paragraph was then imported into the PRAAT software, and a textgrid tier created for ToBI labeling. All files were coded by two experienced ToBI labelers.

The ToBI labels were then subjected to frequency counts and ANOVA analysis to characterize and compare the intonation patterns of native speakers and Mandarin-speaking immigrants with varying AoAs. The following three analyses were conducted: 1) frequency counts and inferential statistical analysis of the prosodic groupings (i.e. intermediate and intonational phrases); 2) frequency counts and inferential statistical analysis of the pitch accents, phrase accents, and boundary tones by type; and 3) dominant pattern analysis.

The first analysis focused on the patterns of prosodic grouping, specifically, the frequencies of Intermediate Phrases (ip) and Intonational Phrases (IP). According to the AM model of English intonation (Beckman & Pierrehumbert, 1986), an ip is signaled by a phrase accent (H- or L-) whereas an IP, which can include one or more ips, is signaled by a boundary tones (H% or L%). The measure of the frequency of the ips was thus the sum of the two types of phrase accents (H- and L-), and the measure of the frequency of the IPs was the sum of the two types of boundary tones (H% and L%). Because of the unequal numbers of participants for each group, the group means, rather than the group sums, were adapted as the measure. To examine the age effect and the specific differences among the groups, one-way ANOVAs and Tukey's Post-Hoc tests were performed. The rationale for the analysis was to examine participants' knowledge of English prosodic structure.

The second analysis investigated the tones, i.e. the pitch accents, phrase accents, and boundary tones, by type. The same analytic techniques (frequency counts, group means, one-way ANOVA and Tukey's Post-Hoc tests) from the first analysis were conducted on all the tone types within each tone category. Since an IP is larger than an ip, an IP boundary tone is always preceded by an ip phrase accent, creating four tonal contours (i.e., L-L%, L-H%, H-L%, H-H%) at the end of an IP. Thus, the second analysis was performed on the nine pitch accent types, two phrase accent types, and four boundary contour types. Similar to the first analysis, this analysis assessed participants' knowledge of English intonation and prosodic structures. Additionally, it examined their understanding of the relationship between semantic and pragmatic meaning and prosody.

Turning now to the dominant pattern analysis, a study conducted by Ross, Ostendorf, & Shattuck-Hufnagel (1992) found variations in pitch accent placements among several native speakers who all read the same news story, suggesting the

difficulties of establishing native speaker standards for prosody. The current findings attested to the study by Ross and colleagues, and demonstrated the variations among native speakers' intonation patterns. Given the nature of intonation variations among native speakers and presumably non-native speakers as well, the current study sought to establish the "dominant patterns," i.e., common patterns of intonation produced by the majority of the speakers within each group, instead of one absolute intonation pattern. A dominant pattern was defined in the current study as the pattern produced by at least half of the participants within each group. Specifically, a dominant pattern for the NS group (N=10) would be the pattern produced by at least five participants within the group and for the Mandarin groups (N=7) would be the pattern produced by at least four participants in each group. Every word was assigned a dominant pattern within each group and dominant tone patterns were then analyzed sentence by sentence. Specifically, the NS dominant tone patterns were first described, and then compared with the three learner groups. The analysis served to describe the differences in the intonation patterns by AoA factor as well as characterize the non-native deviations from native prosody.

1. Prosodic Groupings

The mean frequencies of intermediate phrases and intonational phrases were submitted to one-way ANOVAs, which revealed a significant group effect only for the frequency of Intermediate Phrases, $F(3, 24)=3.14$, $p<.05$. Tukey's tests comparing the learner groups with *NS* further revealed that *Adult Arrivals* produced significantly more intermediate phrases than *NS* ($p<.05$). However, no significant differences were observed among the three learner groups for the frequency of either intermediate phrases or intonational phrases.

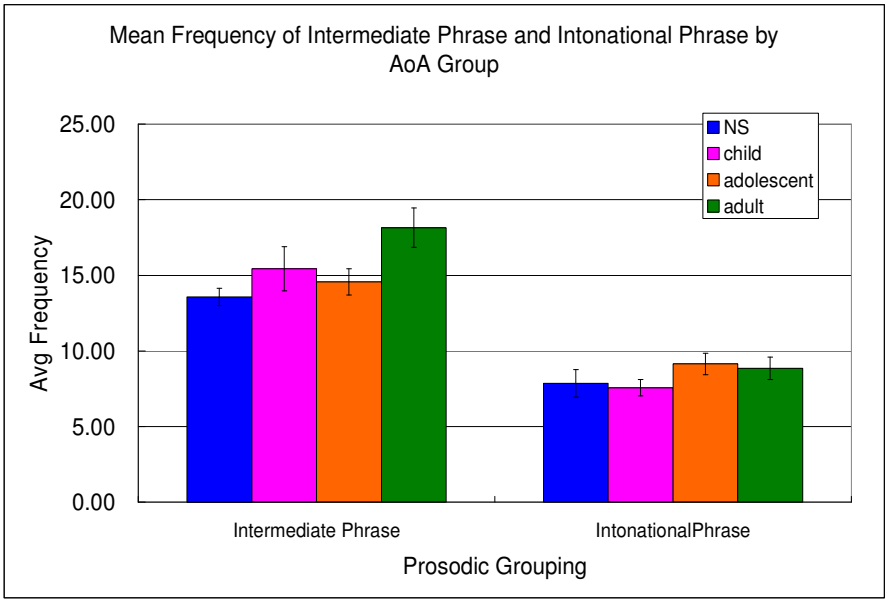

FIGURE 3. Group mean frequencies for intermediate phrases and intonational phrases (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

2. Pitch Accents, Phrase Accents, and Boundary Tones by Type

The mean frequency of the nine pitch accent types were submitted to one-way ANOVAs, which revealed a significant group effect only for the frequency of high tones (H*), $F(3, 24)=5.65$, $p<.000$. Tukey's tests comparing learners groups with NS showed that *Adult Arrivals* produced significantly more high tones than *NS* ($p<.000$). *Child* and *Adolescent Arrivals* were not significantly different from *NS* in the frequency of all pitch accents. Comparing among the three learner groups, *Adult Arrivals* were also found to produced more high tones than *Adolescent Arrivals* ($p<.05$).

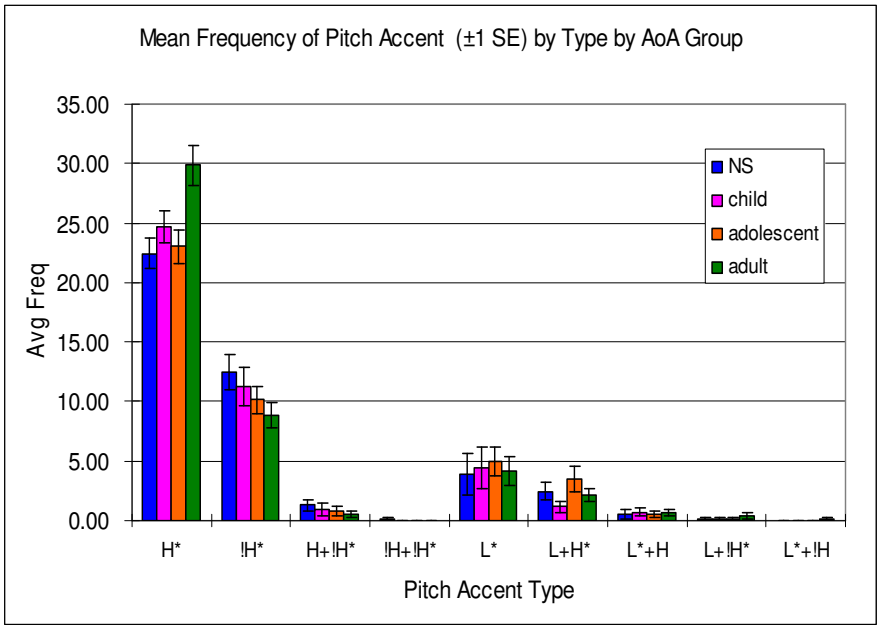

FIGURE 4. Group mean frequencies for pitch accents by type (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

For the two phrase accents (H-, L-) and the four boundary contours (H-L%, H-H%, L-H%, L-L%), there were no significant group effects. However, there were some observed trends of incrementing deviations from *NS* as the age of arrival increased (see Figure 5 and 6).

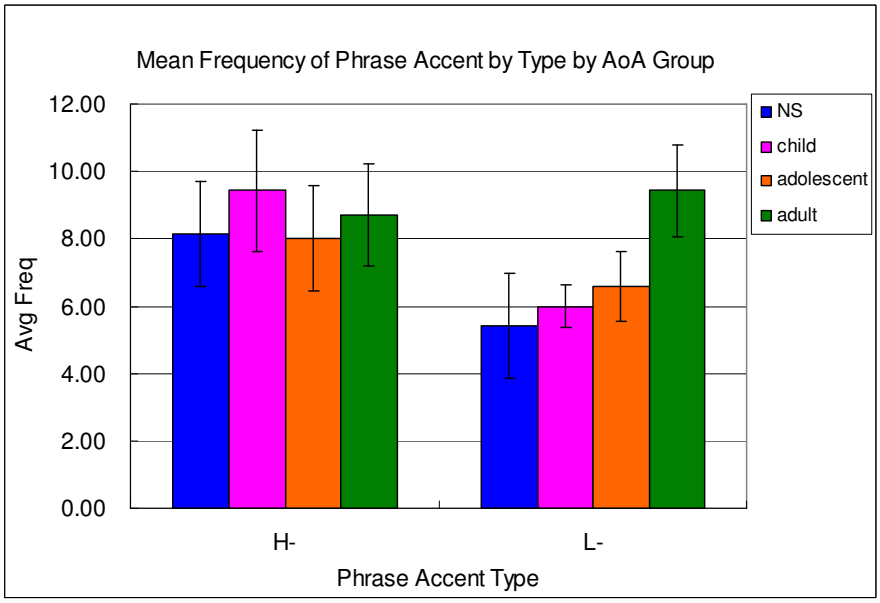

FIGURE 5. Group mean frequencies for phrase accents by type (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

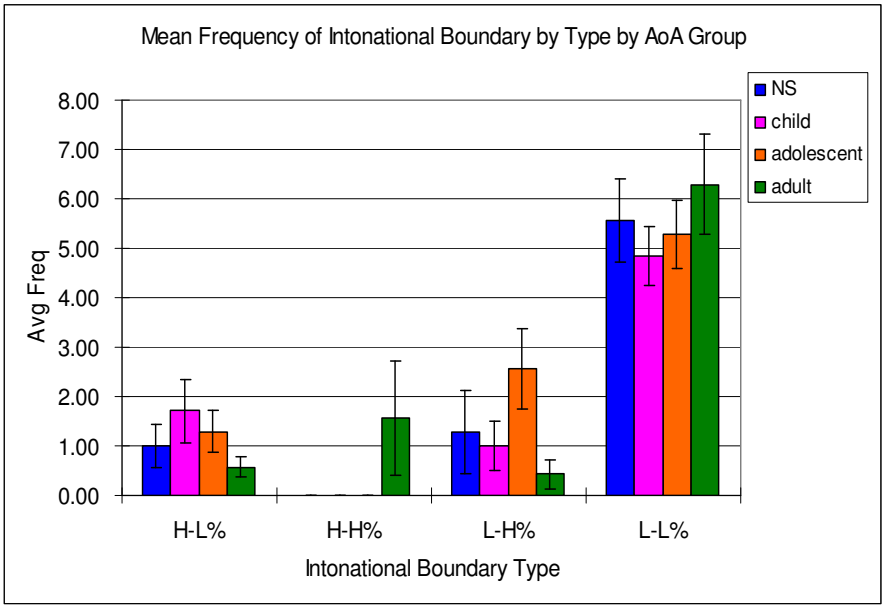

FIGURE 6. Group mean frequencies for boundary contours by type (±1 Standard Error) for the Native Speaker group (NS) and the three learner groups (Child, Adolescent, and Adult Arrivals).

3. Dominant Pattern Analysis

The following section will present the dominant pattern findings sentence by sentence.

- *Please call Stella.*

For this declarative and imperative sentence, the NS group reached high consensuses for all the words (See Table 2). Child Arrivals also produced the same dominant patterns with a high consensus. Adolescent and Adult Arrivals produced the same patterns for the first two words *please* and *call*, but neither group reached a consensus for the last word *Stella*, indicating more variations within the two groups.

- *Ask her to bring these things with her from the store.*

The NS group reached consensuses for almost all the words within the sentence except for two: *bring* and *her* (See Table 3). Child Arrivals were very similar to NS except for a few words. There were also more variations within the Adolescent and Adult Arrivals.

- *Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob.*

For the first and second noun phrases (*six spoons of fresh snow peas, five thick slabs of blue cheese*), the NS group agreed on all words except for *peas, slabs,* and *cheese* (See Table 4). Child Arrivals were similar to the NS group in the frequency of the dominant patterns as well as the dominant patterns themselves. There were more variations among the Adolescent and Adult Arrivals, particularly the Adult Arrivals who

diverged most from the rest of the sample. The Adult Arrivals did not agree on half of the words, and unlike the other groups, they placed a phrase break more often. Furthermore, some of the breaks were placed at an inappropriate or less-ideal position (e.g., phrasal breaks after the preposition *of* are not expected, especially for a reading task in the current study).

For the third noun phrase (*and maybe a snack for her brother Bob)*, the NS group agreed on all words except for *snack* and *Bob* (See Table 5). Child Arrivals were identical to the NS group except for one place: they agreed on a low tone (L*) L-L% pattern for the proper noun *Bob* while the NS group reached no consensus. Adolescent Arrivals were exactly the same as the Child Arrivals both in the frequency of dominant patterns and the dominant patterns themselves. Adult Arrivals also produced similar patterns except for the last two content words, *brother Bob*. However, though not shown in Table 5, some of the Adult Arrivals produced an illegal phrasing by placing a phrase break after *brother* in *brother Bob*.

- *We also need a small plastic snake and a big toy frog for the kids.*

There were some variations within the NS group. They did not agree on several content words: *need, plastic snake,* and *kids.* (See Table 6 & 7). However, they agreed on no pitch accent for all the function words (*we, a, and, a, for, the*). They also agreed on a high tone (H*) for both the adverb *also* and the adjective *small* in the first part of the sentence, and a H* H* !H* H- sequence for the noun phrase *big toy frog*.

There were even more variations within the Child Arrivals, who, in addition to the four non-consensus words that overlapped with the NS non-consensus words, could not agree either on the adverb *also* and the noun phrase *toy frog*. Otherwise their dominant patterns were the same as the NS group. Adolescent Arrivals produced slightly higher consensus patterns than the NS and Child groups, but their dominant patterns patterned closely to those of the two groups. Finally, Adult Arrivals performed similarly to the other three groups in both the frequency of non-consensus words and the dominant patterns for this sentence.

- *She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.*

For the first clause (*she can scoop these things into three red bags*), the NS group agreed on the intonation patterns for almost all but two content words: *things* and *bags* (See Table 8). Child Arrivals, interestingly, did not reach consensus for more than half of the words. As a matter of fact, they were the group with the fewest agreed-upon patterns, suggesting a fair amount of intonation variations within the group for this particular clause. Adolescent and Adult Arrivals were similar in the number of words without

dominant patterns, and those words overlapped with the NS non-consensus words as well. The dominant patterns were also the same between the two groups. To illustrate, they both placed a high tone and a L- phrasal break on the preposition *into* when NS assigned none to the word.

For the second clause *(and we will go meet her Wednesday at the train station),* the NS group reached consensus on all but two words: *her* and *Wednesday* (See Table 9). Child Arrivals agreed on all words except for the noun *Wednesday.* The dominant patterns were similar to those of the NS group except for the two words *we will*, for which the Child Arrivals placed a high tone on the pronoun *we* and a L- phrase accent on the modal verb *will* whereas the NS group assigned none to either word. Unlike the NS group who did not agree on the intonation for the pronoun *her,* the Child Arrivals predominantly produced a high tone with a H- phrase accent.

The Adolescent and Adult Arrivals, on the other hand, were similar to the Child and NS groups among the agreed-upon words. However, there were more variations in the Adolescent and Adult Arrivals as suggested by the larger number of non-consensus words. Take the compound noun *train station* for example, the NS dominant patterns showed that only the first word *train* should receive a high tone pitch accent. In contrast, both Adolescent and Adult Arrivals placed an emphasis on the second word as well, and in a variety of ways (i.e. different types of pitch accents).

Table 2. Dominant Tone Patterns and Numbers of Agreements by Word by Group[4]

	Please	**call**	**Stella**
NS (10)	H*(10)	!H* (9)	!H* L-L% (8)
Child (7)	H* (7)	!H* (5)	!H* L-L% (5)
Adosnt.(7)	H* (6)	!H* (4)	None
Adult (7)	H* (7)	!H* (6)	None

Table 3. Dominant Tone Patterns by Word by Group (cont.)

	Ask	**her**	**to**	**bring**	**these**	**things**	**with**	**her**	**from**	**the**	**store**
NS (10)	0 (8)	H* (8)	0 (10)	None	H* (7)	!H* H- (5)	0 (7)[a]	None[b]	0(10)	0(10)	L+H* L-L%(5)
Child (7)	0 (5)	H* (5)	0 (7)	None	0 (4)	None	0 (5)	None	0 (7)	0 (7)	H* L-L% (5)
Adosnt.(7)	None	None	0 (4)	None	0 (4)	None	0 (6)	None	0(7)	0(7)	L+H* L-L%(4)
Adult (7)	H* (4)	None	0 (4)	None	H* (4)	None	0 (4)	None	0 (5)	0 (7)	None

a. b. One NS participant skipped reading the word.

Table 4. Dominant Tone Patterns by Word by Group (cont.)

	Six	**spoons**	**of**	**fresh**	**snow**	**peas**	**Five**	**thick**	**slabs**	**of**	**blue**	**cheese**
NS (10)	H* (10)	!H* H- (6)	0 (9)	H* (9)	!H* (8)	None	H* H- (5)	H* (10)	None	0 (6)	H* (10)	None
Child (7)	H* (7)	!H* H- (5)	0 (6)	None	H* (4)	None	H* (5)	H* (4)	!H* (4)	None	H* (5)	None
Adosnt.(7)	H* (6)	None	L-(4)	H* (5)	H* (5)	None	H* (6)	H* (4)	None	None	H* (7)	None
Adult (7)	H* (6)	None	L- (4)	None	None	None	H* (6)	H* (5)	None	L- (4)	H* (4)	None

[4] * The number of agreements, i.e. the number of participants in each group who produced the dominant tone patterns, was included in the parenthesis.
* 0= no tones assigned; None= no dominant patterns established

Table 5. Dominant Tone Patterns by Word by Group (cont.)

	and	**maybe**	**a**	**snack**	**for**	**her**	**brother**	**Bob**
NS (10)	0 (10)	H* (8)	0 (10)	None	0 (9)	0 (8)	H* (6)	None
Child (7)	0 (7)	H* (5)	0 (7)	None	0 (7)	0(4)	H* (4)	L* L-L% (5)
Adosnt.(7)	0(7)	H* (5)	0 (7)	None	0 (7)	0(7)	H* (4)	L* L-L% (5)
Adult (7)	0 (6)	H* (4)	0 (6)	None	0 (6)	0 (4)	None	None

Table 6. Dominant Tone Patterns by Word by Group (cont.)

	We	**also**	**need**	**a**	**small**	**plastic**	**snake**
NS (10)	0(10)	H* (6)	None	0(7)	H* (8)	None	None
Child (7)	0(6)	None	None	0(7)	H* (4)	None	None
Adosnt.(7)	0(6)	H* (7)	!H* H- (4)	0(5)	H* (4)	None	None
Adult (7)	0(6)	H* (6)	None	0(6)	H* (4)	H* (5)	None

Table 7. Dominant Tone Patterns by Word by Group (cont.)

	and	**a**	**big**	**toy**	**frog**	**for**	**the**	**kids**
NS (10)	0(10)	0(10)	H* (8)	H* (5)	!H* H- (5)	0(9)	0(10)	None
Child (7)	0 (7)	0 (7)	H* (4)	None	None	0(6)	0(7)	None
Adosnt.(7)	0 (6)	0 (6)	H* (7)	None	!H* H- (4)	0(6)	0 (7)	L+H* L-L%(4)
Adult (7)	0 (6)	0 (6)	H* (6)	H* (4)	None	0(6)	0(7)	None

Table 8. Dominant Tone Patterns by Word by Group (cont.)

	She	**can**	**scoop**	**these**	**things**	**into**	**three**	**red**	**bags**
NS (10)	H* (6)	L- (6)	H* (8)	H* (6)	None	0 (10)	H* (9)	H* (8)	None
Child (7)	0 (4)	None	H* (4)	None	None	None	H* (6)	None	None
Adosnt.(7)	H* (5)	L- (4)	None	None	None	H* L- (4)	H* (6)	!H* (5)	None
Adult (7)	H* (6)	L- (4)	H* (5)	None	None	H* L- (5)	H* (5)[c]	!H* (4)	None

c. disfluent=1

Table 9. Dominant Tone Patterns by Word by Group (cont.)

	and	**we**	**will**	**go**	**meet**	**her**	**Wednesday**	**at**	**the**	**train**	**station**
NS (10)	0(10)	0(7)	0 (5)	H* (9)	H* (5)	None	None	0 (9)	0 (10)	H* (6)	L-L% (8)
Child (7)	0 (6)	H* (7)	L-(4)	H* (5)	H* (4)	H* H- (5)	None	0 (7)	0 (7)	H* (4)	L-L% (6)
Adosnt.(7)	0(6)	None	None	H* (4)	H* (5)	!H* H- (4)	None	0 (7)	0 (7)	H* (4)	None
Adult (7)	0(7)[d]	H* (7)	L- (5)	None	H* (6)	None	None	0 (5)	0 (5)	H* (5)	None

d. disfluent=1

DISCUSSION

The current study aims to investigate the age effect on the ultimate outcome of second language (L2) prosody. Past research on L2 phonological acquisition has primarily focused on the segmental properties, such as consonant production (e.g. Birdsong, 2007; Flege, Frieda, Walley, & Randazza, 1998; Flege et al, 1995), vowel quality (Flege, MacKay, & Meador, 1999; Piske, Flege, MacKay, & Meador, 2002) or global foreign accent[5] (e.g., Flege et al, 1999), rather than L2 prosody. The present study thus undertakes the research effort to fill the gap and explores the age effect on the supra-segmental level of L2 production with speech rate, low-pass filtering technique, and ToBI, the phonological transcription system of prosody.

The current findings suggest that the acquisition of L2 prosody is significantly influenced by the age effect, but the effect is the strongest on Adult Arrivals. The strength of age of learning constraints also varies depending on the specific aspects of prosody under investigation (See Table 10 for a summary of the findings).

First, the findings of speech rate showed that Adult Arrivals were significantly slower than Native Speakers (NS) and Child Arrivals, but were similar to Adolescent Arrivals. Child and Adolescent Arrivals, on the other hand, were not significantly different from NS.

The second analysis of filtered speech ratings suggested that the prosody production of both Adolescent Arrivals and Adult Arrivals were perceived to deviate from NS. Adult Arrivals were also perceived to have a stronger foreign prosody than the Child Arrivals. However, Adult Arrivals were not significantly different from Adolescent Arrivals, and Child Arrivals were perceived to be similar to NS.

The final set of analysis on prosodic grouping and intonation patterns provided further evidence for as well as specificities of the age effect. The analysis of prosodic organizations showed the notable deviations of Adult Arrivals, who produced significantly more phrasal breaks than NS while Child and Adolescent Arrivals were similar to NS. Adult Arrivals' excessive use of phrasal breaks indicated that they divided the paragraph into smaller chunks instead of properly linking the sentences to create a meaning flow. Furthermore, some of the phrasal breaks were placed at an illegal or inappropriate position, suggesting that Adult Arrivals had trouble in processing the sentence and/or in cuing the information structure of the sentence prosodically when reading aloud.

On the other hand, the analysis on the different types of tones (pitch accents, phrase accents, and boundary tones) revealed that Adult Arrivals produced significantly more high tone (H*) pitch accents than both NS and Adolescent Arrivals. According to Pierrehumbert & Hirschberg (1990), H* is generally used by the speakers to introduce new or contrastive information that is to be added to the speaker and the hearer's "mutual belief space" (p. 220). The inappropriate use of H* by Adult Arrivals suggested that they tended to put more emphasis on words regardless of the information status of the word. The findings showed that the Adult Arrivals have not mastered the relationship between meaning and prosody, and thus defaulted to H* even for words with other legitimate tone

[5] As discussed in the background information section, the measure of global foreign accent involves the evaluation of both segments and prosody. The effect of segments is probably stronger than that of prosody in the perception of foreign accents.

assignments. Finally, although no significant differences were found among the groups for phrase accents, boundary tones, and the rest of pitch accents, some trends of age-related declines with later arrivals deviating more from the NS were observed.

Table 10. Summary Table of the Findings

	DESCRIPTIVE Age Trend[a]	ANOVA Main Effect	POST-HOC COMPARISON Tukey's HSD
Speech Rate			
Speech Ratio (num. of syllable per sec.)	√	√	Adult < NS Adult < Child
Filtered Speech Rating			
Mean Rating from 12 Native-speaker Raters	√	√	Adult < NS; Adolescent < NS Adult < Child
Analysis of Tones and Prosodic Groupings			
Pitch Accent, Phrase Accent, & Boundary Contours by Type			
H*	√	√	Adult < NS Adult < Adolescent
!H*	√		
H+!H*	√		
!H+!H*			
L*			
L+H*			
L*+H			
L+!H*	√		
L*+!H	√		
H-			
L-	√		
H-L%			
H-H%	√		
L-H%			
L-L%			
Prosodic Grouping			
Intermediate Phrase	√	√	Adult < NS
Intonational Phrases	√		
Dominant Pattern Analysis			
Dominant Patterns	√[b]	NA	NA

a. An age trend is checked when the Child Arrival group was found to be more similar to the NS group than the Adult Arrival group in descriptive results (e.g. frequency counts), regardless of the inferential statistical test results.

b. There were more variations (i.e. lower frequency of dominant patterns) in the Adult Arrival group than in the Child or Adolescent Arrival groups.

Turning now to the dominant pattern findings, as suggested by Ross et al (1992), some variations were observed even in native speakers' intonation patterns. However, for all words in the paragraph considered there, the NS group showed about eighty percent agreement. Among the learner groups, Child and Adolescent Arrivals overall pattern similarly to the NS group, and there were notably the most variations within the Adult Arrivals. Not only did Adult Arrivals disagree on most words, but their dominant patterns diverged the most from the NS dominant patterns. They either placed pitch accents and phrase breaks at illegal locations, or placed the wrong type of tones, violating the relationship between prosody and pragmatic and semantic meanings.

Taken together, the current study confirmed the age effect on the acquisition of L2 prosody. Despite similar length of residence, the prosody of Adult Arrivals remained quantitatively and qualitatively distinct from the native speakers. However, the strength of the learning constraints varied among different aspects of prosody. Depending on the specific aspect of prosody under study, Adult Arrivals can also differ considerably from the Child and/or Adolescent Arrivals. For example, Adult Arrivals diverged significantly from NS and Adolescent Arrivals in the frequency of high tones. They also had a significantly lower speech rate and foreign prosody rating than both NS and Child Arrivals, and the most variations in intonation patterns among all groups. But in terms of the frequency of phrase accents, boundary tones, or most of the other pitch accent types, Adults were not significantly different from the other groups. On the other hand, Child Arrivals appeared to be identical to NS, and Adolescent Arrivals generally resembled NS except in the foreign prosody rating and certain dominant intonation patterns.

CONCLUSION & FUTURE DIRECTIONS

To conclude, our exploratory study provided evidence for an advantage in early learning of L2 prosody. Due to constraints of the design with AoA as a categorical rather than a continuous variable, however, it did not delineate a specific timeframe for the prosody acquisition. The current findings thus did not support a predetermined terminus of L2 learning potential at puberty in the traditional "critical period" sense. However, the findings might be interpreted as support for a "sensitive period" that proposes declines in plasticity and increases of variations in L2 outcomes with increasing ages of learning.

To clarify the timeframe of the age constraints on L2 prosody learning, future research should use AoA as a continuous variable and investigate and control variables that possibly confound with AoA, such as the amount of language exposure and years of U.S. schooling (Flege, 2007).

Future research would also greatly benefit from methodological improvements in establishing native speaker baseline patterns as well as in quantifying L2 learners' prosodic deviances from the native speaker patterns. For example, to reduce the variations in native speakers' intonation patterns, stimuli could be constructed with constrained contexts and responses to yield more consistent baseline patterns for comparison. Additionally, exploring or creating novel analytical techniques to characterize prosody, such as using ToBI model to analyze phonological tonal categories as in the present study, would also contribute to advancing methodology in L2 prosody research. Assessing the phonetic realizations of the tonal categories can also contribute to advancing the analysis of L2 prosody (see Ueyama & Jun, 1998 for the comparison

among groups based on phonetic realizations of H- tone in English). We are currently exploring ways to quantify different realizations of tonal categories.

Finally, it will be equally beneficial to research the sources and causes of the perception of accentedness to understand the relative importance of segmental and suprasegmental features in foreign accent (e.g., Vicenik & Sundara, 2008) as well as the contributions of various suprasegmental components in foreign prosody. The current study serves as a preliminary effort in the investigation of the age constraints in ultimate L2 prosody attainment. Further research in the above-mentioned directions is clearly needed before a comprehensive picture can be unveiled.

ACKNOWLEDGEMENT

We are grateful to James Flege, Peggy Li, and members of the UCLA Phonetics Laboratory for their assistance and feedback at various stages of the study. We also thank Marissa Tse, Alex Wang, Tammy Ko, Tia Fang, Tiffany Tseng and Eric Williams for their help with data collection, speech segmentation, and ToBI labeling. Special thanks to all the participants in the study for their time and patience.

The study was funded by the dissertation grants from the Language Learning Journal Dissertation Grant Program and the Chiang Ching-Kuo Foundation for International Scholarly Exchange to the first author, and the UCLA Faculty Research Grant to the second author.

REFERENCE

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529–555.

Asher, J., & Garcia, R. (1969). The optimal age to learn a foreign language. *Modern Language Journal, 53,* 334–341.

Beckman, M. & Ayers-Elam, G.(1994). *Guidelines for ToBI labeling.* Retrieved May 15, 2008, from Ohio University Linguistics Department Website: http://www.ling.ohiostate.edu/research/phonetics/E_ToBI/singer_tobi.html

Beckman, M. E., & Hirschberg, J. (1994). The ToBI Annotation Conventions. Unpublished manuscript, Ohio State University.

Beckman, M. & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255-310.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins.

Boersma, Paul & Weenink, David (2008). Praat: doing phonetics by computer (Version 5.0.45) [Computer program]. Retrieved December 29, 2008, from http://www.praat.org/

Bruce, G. (1977) *Swedish word accents in sentence perspective*. Lund: Gleerup. 1982. Developing the Swedish intonation model. *Working Papers*, Department of Linguistics and Phonetics, University of Lund 22: 51-116.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice.* Amsterdam: Benjamins.

Flege, J. E. (2008). Give input a chance. In T. Piske & M. Young-Scholten. (Eds.), *Input Matters in SLA.* Buffalo: Multilingual Matters.

Flege, J., Frieda, E., Walley, A., & Randazza, L. (1998). Lexical Factors and Segmental Accuracy in Second-language Speech Production. *Studies in Second Language Acquisition, 20*, 155-188.

Flege, J. E., MacKay, I., & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of Acoustic Society of America, 106*, 2973–2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995a). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16,* 1–26.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, *41*, 78–104.

Golinkoff, R. M (1983). *The Transition from Prelinguistic to Linguistic Communication.* Hillsdale, NJ: Lawrence Erlbaum.

Guion, S.G., Flege, J.E.; Liu, S. & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21*, 205-228.

Hirschberg, J. & Pierrehumbert, J. (1986). Intonational structuring of discourse. *Proceedings of the 24th meeting of the Association for Computational Linguistics*, New York, 126-144.

Huang, B. H. (forthcoming). The Earlier the Better? Age Effect on Second Language Phonology and Morphosyntax. To appear in the *Proceedings of the International Conference on Applied Linguistics*. Chiayi, Taiwan.

Jun, S-A, & Oh, M. (2000) Acquisition of Second Language Intonation, *Proceedings of International Conference on Spoken Language Processing, 4*, 76-79.

Kjelgaard, D. & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*, 153-194.

Ladd, D. R. (1996). *Intonational Phonology.* New York, NY. Cambridge University Press.

Lehiste, I (1970). *Suprasegmentals.* Cambridge: The MIT Press

Li, C. N. & Thompson, S.A, (1977). The Acquisition of Tone in Mandarin-speaking Children. *Journal of Child Language, 4*, 185-199.

Liberman, M. (1975). *The intonational system of English*. Doctoral dissertation. MIT. Cambridge, MA. Distributed 1978 by IULC.

MacKay, I.R.A., Flege, J.E. & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, *27,* 157–183.

Magen, H.(1998). The perception of foreign-accented speech. *Journal of Phonetics, 26,* 381-400.

MATLAB version 6.5.1. Natick, Massachusetts: The MathWorks Inc., 2003.

Millotte, S., Wales, R. & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes, 22*, 898-909.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition, 21*, 81-108.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17*, 17–34.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97.

Neufeld, G. G. (1978). On the acquisition of prosodic and articulatory features in adult language learning. *Canadian Modern Language Review, 34,* 168-194.

Oyama, S. (1976) A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*, 261–283.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

Pierrehumbert, J. & Hirschberg, J. (1990). The Meaning of Intonation in the interpretation of Discourse, In P. Cohen, J. Morgan, & M. Pollack, (Eds), *Intentions in Communication,* (pp. 271-311). Cambridge MA. MIT Press,

Piller, I. (2002). Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics, 6* (2), 179-208.

Piske, T., Flege, J., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*, 49-71.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation, *Journal of Acoustical Society of America, 90*(6), 2956-2970.

Ross, K., Ostendorf, M, & Shattuck-Hufnagel (1992). Factors affecting pitch accent placement. *Proceedings of the International Conference on Spoken Language Processing,* 365-368.

Scovel, T. (1988). *A time to speak: a psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.

Shi, R. & Moisan, A. (2008). *Prosodic Cues to Noun and Verb Categories in Infant-Directed Speech.* In H. Chan, H. Jacob, and E. Kapia (Eds), Proceedings of the 32nd Boston University Conference on Language Development (pp. 450-461). Somerville, MA: Cascadilla Press.

Speech Accent Archive website: http://accent.gmu.edu/

Trofimovich, P., & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Ueyama, M. & Jun, S.-A. (1998). Focus realization in Japanese English and Korean English intonation, *Japanese and Korean Linguistics, 7*, 629-645.

Vicenik, C. & Sundara, M. (2008). Role of segmental and intonational cues in dialect discrimination. Poster Presented at *Acoustics'08 the 155th meeting of the Acoustical Society of America*, Paris, France.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition, 20* (1), 1–25.

APPENDIX I

Paragraph Stimuli

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Source: The Speech Archive Website