



The Effects of ESL Immersion and Proficiency on Learners' Pronunciation Development

Maria Kostromitina and Okim Kang*

Northern Arizona University, Flagstaff, AZ, United States

Despite the efforts of existing studies in the domain of L2 phonology to examine ESL learners' pronunciation development, little research has comprehensively demonstrated ESL learners' pronunciation improvement in academic immersion contexts. Similarly, few studies have focused on learners' proficiency levels linked to their developmental success. The current exploratory study investigated the changes of learners' pronunciation constructs as a result of their ESL program. Seventy-five newly arrived ESL students (25 in each proficiency; beginner, intermediate, and advanced) enrolled in an Intensive English Program in the United States provided their speech responses (to the placement and exit tests from the program). One hundred fifty speaking samples were linguistically analyzed for the following suprasegmental features: fluency (speech rates and pauses) and prosody (prominence and pitch range). Segmental features were analyzed by employing a functional load approach with randomly selected 90 speech files. Findings revealed different developmental patterns among phonological features and proficiency levels; that is, the upper-level learners improved more in fluency and prominence than the lower-level learners. Segmental changes were minimal, suggesting that both high functional and low functional load sounds involve a complex process in learning. Overall findings provide important implications for ESL curriculum planning and development: 1) intonation acquisition can be difficult; 2) skill improvement differs by proficiency level; and 3) level-specific curriculum may be needed.

Keywords: ESL immersion, L2 pronunciation, L2 proficiency, functional load, fluency

OPEN ACCESS

Edited by:

Andrew Sewell,
Lingnan University, China

Reviewed by:

Paul Stapleton,
Other, Toronto, ON, Canada
Dustin Crowther,
University of Hawaii at Manoa,
United States

*Correspondence:

Okim Kang
okim.kang@nau.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 30 November 2020

Accepted: 19 February 2021

Published: 22 April 2021

Citation:

Kostromitina M and Kang O (2021) The
Effects of ESL Immersion
and Proficiency on Learners'
Pronunciation Development.
Front. Commun. 6:636122.
doi: 10.3389/fcomm.2021.636122

INTRODUCTION

Recent developments in the field of second language (L2) speech have instigated a shift of the emphasis on intelligible speech over the unrealistic goal of sounding native-like (Derwing and Munro, 2005; Field, 2005). Following this trend, instead of a complete absence of a foreign accent, second language (L2) English speakers are encouraged to strive for clear and intelligible speech. Previous research on L2 speech has identified that functional load-based segmental deviations can predict speaker intelligibility (Kang et al., 2020) or distinguish proficiency levels (Kang and Moran, 2014). Other studies have demonstrated that suprasegmental features, including fluency, prominence, and intonation, have an even larger role in promoting L2 speakers' intelligibility and comprehensibility in oral communication (Derwing and Rossiter, 2003; Pickering, 2004; Kang et al., 2010).

At the same time, study abroad (SA) experiences or immersion (IM) contexts, especially combined with instruction, are known to be beneficial for L2 learners' pronunciation

development (Stevens, 2002; Lord, 2010). However, this development can happen differently depending on certain learner factors. In a series of longitudinal studies with Mandarin and Slavic ESL learners in Canada, Derwing and colleagues investigated the development of comprehensibility, fluency, and accentedness of L2 speakers as well as the relationship between L1 and L2 fluency (Derwing et al., 2006; Derwing et al., 2008; Derwing et al., 2009). The studies demonstrated differences in the improvement of these L2 pronunciation features between the two groups, emphasizing the role of motivational, cultural, and interactional factors in L2 pronunciation development in immersion (Derwing and Munro, 2013).

Such individual differences can also predict learners' success in developing linguistic competence as a result of L2 immersion. For example, after studying abroad, advanced learners brought more formulaic expressions to communicative contexts whereas beginners attended primarily to meaning (Lafford, 2004). Segalowitz and Freed (2004) found that among Spanish L2 learners, an initial threshold level of basic word recognition and lexical access processing abilities was necessary for a significant oral proficiency development in the immersion context. In their study, the most important gains that learners of Spanish made abroad were in the domain of speech fluency.

Regardless of the empirical evidence about the benefits of immersion experience in language development, however, few studies have comprehensively demonstrated L2 learners' pronunciation improvement, particularly with a focus on their ESL immersion experience through a functional load approach. Similarly, how learners' proficiency is linked to their developmental success in L2 pronunciation has been largely unknown. The present study explores the effects of ESL immersion on learners' pronunciation improvement across different levels of learners' proficiency. The results of the study shed light on pronunciation development of L2 English learners and offer important implications for curriculum design of ESL immersion programs.

REVIEW OF LITERATURE

ESL Immersion and Learners' L2 Proficiency

When discussing linguistic gains in an immersion environment, it is important to understand what immersion entails. Freed et al. (2004) defined immersion as a combination of classroom-based learning with expected outside activities in the at home environment. In an ESL context, immersion is most commonly associated with Intensive English Programs (IEP) at universities. Such programs are different in their duration and are typically designed for learners from various proficiency levels. Beyond IEP classrooms, immersion often includes opportunities for interaction with the native speech community and out-of-class activities.

In light of IEP, research has been focused on identifying the most favorable period in learners' L2 development when they can gain the most from immersion, bringing forward the idea of

“proficiency threshold.” Despite the general consensus about the benefits of immersion, there is little agreement in the existing literature on the most beneficial time for a language learner to immerse in the target language. Attempts have been made to determine the optimal point in learners' proficiency to be immersed for it to result in noticeable L2 development. For example, Kang and Ghanem, (2016) found with the help of a nation-wide self-report survey that the intermediate level was somewhat more beneficial than other levels for immersion programs. Other researchers (e.g., Brecht et al., 1995; Martinsen, 2008; Collentine, 2009; Muñoz and Llanes, 2014) argued that beginning-level language learners demonstrate the greatest amount of improvement in oral and aural communication skills as a result of ESL immersion. In contrast, others (e.g., Davidson, 2010) called for more studies to focus on advanced-level learners to validate the current findings.

Notwithstanding the research discrepancies, low-level proficiency learners seem to benefit from immersion contexts. As Martinsen, (2008) pointed out, true beginners are likely to be slower to progress than learners with at least some experience with the target language indicating “there may be a minimal level of proficiency at which learning abroad is optimal” (p. 506). Additionally, for students' oral abilities to improve, learners need at least a basic level of word recognition and processing abilities (Segalowitz and Freed, 2004). When discussing the proficiency threshold in the context of ESL immersion, however, learners' linguistic competence is indeed a complex construct (Collentine, 2009). How learners' proficiency levels relate to their learning gains in the immersion context needs further validation.

L2 Pronunciation Improvement in the Immersion Context

Pronunciation in an L2 is commonly operationalized as “aspects of oral production of language, including segments, prosody, voice quality, and rate” (Derwing and Munro, 2015, p. 5). However, as Derwing and Munro (2015) note, in interactional situations, pronunciation encompasses broader dimensions of communication that include *accentedness*, or particular patterns of pronunciation that distinguish members of speech communities; *comprehensibility*, or the amount of effort with which a listener understands L2 speech; and *intelligibility*, or the degree to which a message is received as intended by a listener. Another speech dimension related to pronunciation is fluency, or the rate and degree of fluidity of speech, indicated by the absence of pauses or other disfluency markers.

In ESL immersion contexts, L2 pronunciation has received relatively consistent attention in the field of SLA. Overall findings point out that advances in pronunciation over time are a slow or unchanging process (Avello, 2010; Pérez-Vidal et al., 2011). Trofimovich and Baker (2006) examined the changes in English learners' speech rate as a result of their U.S. immersion experience over different time periods (i.e., three months, three years, and ten months) and showed that there was no significant difference in the speech rate among the learners. Despite the length of residency in the U.S., speech

rate did not necessarily change for those learners. In fact, the authors suggested that certain suprasegmental features, including speech rate and tonal peaks, may never be learned to a native-like proficiency (Trofimovich and Baker, 2006).

More recently, Højen (2019) examined effects of short-term immersion (3–10 months) on adult L2 English pronunciation. In the study, native English speakers evaluated the accentedness in speech samples from three experimental groups: a native Danish au pair group with experience living in England, a native Danish control group with no such experience, and a native English reference group. For the experience group, speech samples were recorded twice, before and after the immersion. The results of the study revealed significant improvements when compared before and after the immersion with a great degree of variation. Interestingly, the pronunciation score for the immersion group was significantly correlated with the length of residence in England ($r = 0.61$) suggesting that an immersion of at least five months is needed for noticeable improvement in L2 pronunciation. More longitudinal research can confirm such developmental patterns.

Pronunciation improvement over time in an ESL immersion context can be confounded by learners' first language (L1) background, even though learners' L1 is not the primary focus of the current study. In a longitudinal study over ten months, Derwing et al. (2006) evaluated the progress in Slavic and Chinese Canadian immigrants' accent and fluency over a period of 10 months. Both groups only showed a small improvement in accentedness over time, but the two L1 groups had different results in terms of fluency. The Slavic learners demonstrated significant improvement in fluency, while the Mandarin group did not. Considering these same L1 groups over a two-year period, Derwing et al. (2008) found similar results in their later research. In a more recent study, Derwing and Munro (2013) examined the extent to which the same two groups of learners continued to make progress in the development of oral English skills after finishing their formal ESL training. Similarly to earlier findings, Slavic speakers improved comprehensibility and fluency using English outside of the classroom context while Mandarin speakers showed much less improvement. The authors proposed that MacIntyre's (2007) Willingness To Communicate framework could account for the differences underlying the performance of the two groups including ties to the L1 community, reluctance to initiate conversations, and lack of opportunities to interact in English.

While there is developmental variation among different L1 groups, fluency certainly seems to benefit the most in an immersion context (Freed et al., 2004; Segalowitz and Freed, 2004). Towell (2002) further added that the initial fluency of the learners determined learners' fluency gains; in other words, fluency improved most significantly at the lower levels than at the higher levels. Overall, it seems essential to systematically determine what pronunciation properties underlying comprehensibility, intelligibility, accentedness, and fluency of L2 speech are actually changeable or learnable over time especially in such an immersion context and to what extent these properties may improve as a result of immersion.

Functional Load-Based Segmental Features in L2 Pronunciation

Segmental analyses of pronunciation often involve an examination of deviations from the native baseline or substitution of sounds in L2 speech (e.g., fun spoken like fan, Isaacs and Trofimovich, 2012). Other studies (e.g., Kang and Moran, 2014) have categorized the segmental deviations in terms of their relative weight in predicting listeners' judgments. Not all segmental deviations can have equal effects on listeners' understanding (see also Fayer and Krasinski, 1987) and a more "nuanced approach" is needed (Isaacs and Trofimovich, 2012).

An effort that has been made to categorize segmental deviations is the Functional Load (FL) theory (Catford, 1987; Brown, 1991). Munro and Derwing (2006) further explain that segmental pairs (e.g., pet vs. bet or dis vs. this) are ranked based on factors including the probability that individual members of the minimal pair are valid, the frequency of the minimal pair, and the position of the segmental within a word. Thus, if an L2 speaker substitutes 'them' for 'dem', it is unlikely that their comprehensibility is severely affected in a negative way. In contrast, the /d/-/p/contrast has a high functional load in English meaning (e.g., day-pay). Munro and Derwing (2006) demonstrated that high FL divergences had larger effects on listeners' perceptions of accentedness and comprehensibility of L2 speech than low FL deviations thus providing preliminary support for the theory.

In an L2 assessment study, Kang and Moran (2014) classified segmental deviations according to their FL to determine their effect on oral assessment across four proficiency levels (B1–C2). The analysis of vowel and consonant substitution divergences through the FL approach detected a significant difference across proficiency levels in the high FL deviations. That is, with an increase in learners' proficiency, the amount of high FL deviations dropped significantly. However, changes in low FL deviations were not noticeable across levels.

In a recent study, Suzukida and Saito (2019) re-examined the FL approach to evaluating the effect of segmental divergences on L2 comprehensibility in two experiments with learners in EFL and immersion settings. In the first experiment, the speech of Japanese learners of English in EFL settings was assessed in terms of perceived comprehensibility by L1 English raters. The second experiment was slightly different with the speakers being Japanese learners of English with immersion experience. Their findings also showed that only high FL consonant substitutions negatively affected native listeners' comprehensibility judgments in both experiments significantly. Importantly, high FL consonant substitutions impeded raters' comprehensibility regardless of task conditions. Kang et al., 2020 recent study also confirmed that divergences in high FL vowels and consonants strongly predicted listener comprehension and intelligibility scores.

It is evident that FL-based segmental pronunciation features play a critical role in listener judgments and perceptions of L2 speech. However, while some studies analyzed the differences in FL deviations in speech of L2 learners from different proficiencies

(e.g., Kang and Moran, 2014), there have been only a few studies that investigated the development of segmental features over time. In particular, Kang et al., 2021 (in press) recent study examined how EFL learners developed their speaking skills by analyzing fifty-two EFL learners' IELTS spoken responses over the period of three months. The study comprehensively analyzed segmental and suprasegmental pronunciation features but did not find significant improvements in segmentals. The authors called for further research.

Suprasegmental Features of L2 Pronunciation

An increasing number of studies have addressed the importance of suprasegmentals, such as fluency, stress, and intonation, in listeners' judgments of accentedness and comprehensibility of L2 speakers (Munro and Derwing, 2001; Isaacs, 2008; Kang et al., 2010). Research indicates that fluency, characterized by the speaking rate, number and length of pauses, and repair fluency, is linked to listeners' comprehension of speech and the overall evaluation of speakers' oral proficiency (Derwing et al., 2004; Tavakoli and Skehan, 2005; Iwashita et al., 2008). With regard to accentedness, Trofimovich and Baker (2006) indicated that accentedness ratings given by L1 English speakers to Korean learners of English were higher when the L2 speech was faster. In fact, several studies have suggested certain speaking rate thresholds, which make the perception of L2 speech more comprehensible and less accented (Isaacs, 2008; Munro and Derwing, 2001). Kang et al. (2020) recently reported that temporal fluency measures predicted listener comprehension and intelligibility scores. Other studies showed that pause frequency and duration affected accentedness and comprehensibility ratings (Trofimovich and Baker, 2006; Kang et al., 2010).

Other suprasegmental features that have been found especially indicative of L2 pronunciation development include nuclear stress and pitch range. It has been established that placing incorrect sentence stress or emphasizing every word in a run, regardless of its function or importance to the communicative purpose, can negatively affect listeners' comprehension (Juffs, 1990; Wennerstrom, 2000; Field, 2005). Similarly, pitch range that is too narrow can considerably diminish L1 listeners' comprehension of L2 speech (Pickering, 2001) and cause misunderstandings (Kang, 2012). Moreover, in conjunction with accentedness ratings and suprasegmentals, Kang (2010) found that pitch range alone explained 24% of the variance in accentedness ratings, with narrow pitch range being associated with stronger accents. Thus, the particular suprasegmental features that were measured in this study (fluency, stress, and pitch range) reflect previous research that has systematically shown the importance of these features for the perceptions of L2 speech accentedness and comprehensibility making them crucial in L2 pronunciation.

Similar to that of segmental research, however, the development of suprasegmental features has rarely been studied, especially from a longitudinal perspective. Kang et al., 2021 (in press) demonstrated EFL learners' fluency and

intonation changes over 12 weeks, but their findings are very limited. Generally, research that has comprehensively examined the production of both segmental and suprasegmental features, especially pertaining to their development over time, is scarce. In addition, little is known about the way learners at different proficiency levels develop pronunciation skills in an immersion context. In an attempt to fill these gaps, the present exploratory study systematically examines the interplay of proficiency and immersion on the changes in L2 learners' FL-based segmental and suprasegmental pronunciation. The present study addresses the following research question:

To what extent does ESL learners' pronunciation develop as a result of one-semester long ESL immersion across the proficiency levels?

- 1) In terms of functional load-based segmentals in terms of suprasegmentals (fluency, pitch, sentence stress).

METHODS

Participants

The study recruited seventy-five ESL students enrolled in listening and speaking courses in an Intensive English Program (IEP) at a southwestern university in the United States. There were 25 participants from each of the three proficiency levels in the program: beginning, intermediate, and advanced. These proficiency levels were determined through an in-house placement test, which mimicked the standardized iBT TOEFL test of English proficiency. Level 1, or beginner, corresponds to scores below 15 of the TOEFL iBT; Level 3, intermediate, corresponds to scores between 32–44; and Level 5, upper-intermediate, corresponds to scores between 57–69. The first language of 59 speakers was Arabic while the remaining 16 participants were native speakers of Chinese. Most of them (90% of the participants) just arrived in the U.S. and started the IEP program for the first time while the remaining 10% arrived slightly earlier (1–2 weeks).

Speaking Tasks and Collection of Speech Samples

There were two stages of speech sample collection. The first collection of the samples took place during the first week of classes (pre-immersion) as a placement test to determine the appropriate level for the learners in the IEP listening and speaking class. The second speech sample collection happened during week 15 (post-immersion) when the learners completed an exit test to demonstrate that they successfully finished the course and were ready to move up to the next level in the program or graduate from the IEP.

During both times of the speech collection, the participants completed the same speaking task. The task consisted of an oral prompt that asked the participants to speak on the following topic: Some students prefer university in their home country, while others prefer studying abroad. What do you prefer? Give

TABLE 1 | Summary of speakers response length before and after the immersion.

	Pre-immersion (time 1)			Post-immersion (time 2)		
	Beginner	Intermediate	Advanced	Beginner	Intermediate	Advanced
<i>N</i>	25	25	25	25	25	25
<i>M</i> (sec)	45.68	63.25	88.83	58.68	62.02	79.80
<i>SD</i>	21.12	13.74	18.29	18.05	12.12	24.62

TABLE 2 | Pronunciation feature analysis and measures.

Categories	Feature	Description
Segmental measures	High FL substitutions	This measure calculates the number of high functional load consonant and vowel substitutions in word initial, word medial, or word final positions. e.g., “day”–“bay”; “cat”–“cot”
	Low FL substitutions	This measure calculated the number of low functional load consonant and vowel substitutions in word initial, word medial, or word final positions. e.g., “this”–“zis”; “walking”–“wolking”
Fluency measures	Syllables per second	This is a measure of the mean number of syllables produced per second, calculated as the total number of syllables divided by the total length of the speech sample
	Number of silent pauses per second	This measure is the number of silent pauses per second, calculated as the total number of pauses (over 0.1 s) divided by the total length of the speech sample.
	Number of hesitation markers per second	This measure is calculated as the total number of filled pauses divided by the total length of the speech sample. Filled pauses include hesitation markers or fillers such as <i>uh</i> or <i>um</i> but do not include repetitions, restarts, or repairs.
Stress Measure	Space	This measure calculates the proportion of prominent words to the total number of words.
Intonation measure	Overall pitch range	This measure calculates the pitch range of the sample based on the F_0 minimum and maximum frequency point on all prominent syllables within the speech sample.

reasons and examples to support your opinion. The task was designed to elicit monologic speech samples from the speakers. While presenting the same task to the participants before and after the immersion could present a potential risk of task familiarity effect, the importance of increased comparability of the speech samples outweighed this concern. Each participant was permitted to ask questions regarding the content of the prompt and any unknown vocabulary and given 1.5 min to prepare their response. The participants were then advised to speak for about 1–2 min on the topic and were recorded the entire time. The final dataset included pre- and post-immersion speech samples from the 75 participants. There were a total of 150 speech files varying from 30 s to 2 min each; that is, each participant contributed two sound files to the dataset. **Table 1** below presents the summary of the descriptive statistics (*M* and *SD*) of the speech samples before and after the immersion.

Speech Analysis

To analyze the development patterns of the pronunciation of segmentals and suprasegmentals in the speech samples, the 150 sound files were first transcribed by three trained coders. Next, FL deviations in a randomly sampled subset of 90 sound files were analyzed, calculated, and averaged per minute. It was deemed appropriate to subsample the files for the FL analysis as this process required a much more meticulous examination of the speech samples; thus, the difference in the sample size between segmental and suprasegmental analyses happened due to the labor intensiveness of speech analysis involved in different types of speech features. This subset consisted of 45 speech samples collected before the immersion and 45 speech samples collected

after the immersion produced by the same speakers (15 speech samples per proficiency level in both cases). The FL deviations were categorized into two groups: high FL divergences and low FL consonant deviations (Catford, 1987; Kang et al., 2020). Following Munro and Derwing (2006) and Kang and Moran (2014), we considered the substitutions that ranked between 51 and 100% in Catford’s framework as high FL divergences, and those below were regarded as low. To identify these deviations, the coders listened and transcribed the speech files noting all the instances when a speaker’s pronunciation deviated from Standard American English (SAE). The coders then used Catford’s FL framework to assign a functional load value (percentage) to the deviations. Additionally, we analyzed the suprasegmental measures for all of the 150 speech samples.

The complete list of measures is presented in **Table 2**. The specific segmental measures selected for the pronunciation analysis have been found to differ distinctively across proficiency levels (e.g., Kang and Moran, 2014). The suprasegmental measures chosen for this analysis are also grounded in the aforementioned previous research that emphasized their weight for comprehensibility and accentedness judgments of L2 speech and for L2 pronunciation overall (e.g., Pickering, 2001; Kang, 2010; Kang et al., 2010). In accordance with Kang (2010), the runs in the speech samples were operationalized as stretches of undisturbed speech delimited by pauses of 0.1 or longer assuming that this pause cut-off would be meaningful in L2 speech.

To prepare the speech samples for analysis, they were converted into .wav format with the help of Audacity, a free audio software (Audacity Team, 2020) and transcribed using standard conventions. Then, three trained coders analyzed the

FL-based deviations in the samples. The coders were unaware of the proficiency levels of the speakers or any other identifying information. One of the coders calculated inter-coder reliability post hoc by re-analyzing 10% of the speech and reaching over 85% agreement with each coder. All remaining differences were later discussed and resolved reaching 100% agreement among the raters. As for the suprasegmental features, Praat (Boersma and Weenink, 2020) was used to conduct temporal and acoustic analyses. Spectrograms with extracted pitch contours were used to identify the prominent syllables in runs and consequently measure the pitch on prominent syllables as well as the length and number of silent and filled pauses. Two coders performed analysis of the suprasegmental features after reaching an agreement of 90% or higher on a subset of data for all variables.

Statistical Analysis

The main research question in the present study addressed the change in FL-based segmental and suprasegmental pronunciation features of beginner, intermediate, and advanced ESL learners as a result of one semester-long immersion. To answer this research question, seven linear mixed effect models (LMEM) were performed with each of the seven pronunciation features in **Table 1** as dependent variables and proficiency and time pre-/post-immersion as independent variables. The three pre-determined proficiency levels in the study and the time of speech sample collection (Time 1 and Time 2) were entered as fixed factors. Participants and their L1 backgrounds were entered as the random factors. Linear mixed effects modeling was considered appropriate for the current analysis since it allowed to investigate the effects of immersion, proficiency, and their interaction on pronunciation features while controlling for the participant and L1 (Arabic and Chinese) background factors. All statistical procedures in the study were completed with the help of *R* (ver. 4.0.2), a free statistical environment (R Core Team, 2020). Before fitting the models onto the data, seven scatterplots were created to examine the data for violations of normality, linearity, and homoscedasticity. After winsorization of the outliers, the data met the assumptions.

RESULTS

The goal of the present study was to examine the change in segmental and suprasegmental features in speech of beginner, intermediate, and advanced English learners as a result of ESL immersion. The following section provides a detailed description of the results of the mixed effects models fitted on the data using one dependent variable at a time. First, the results of the segmental analysis are given with regards to the high and low FL deviations in the speech samples. Both types of FL deviations each included consonant and vowel substitutions. Then, we present the analysis summary of the five suprasegmental features that represented fluency (speaking rate, silent pauses, and filled pauses), stress (space), and intonation (pitch range) in our study.

Development of FL-Based Segmentals

The first two LMEMs fitted on the data focused on High and Low FL substitutions in the speech samples. Importantly, to address

the development of FL-based segmentals, high FL vowels and consonants were merged as well as low FL vowels and consonants. While we initially attempted to analyze them separately, the analysis revealed similar deviation patterns among the high FL vowels and consonants and low FL vowels and consonants. To allow for a more robust sample, the categories were combined into two groups: high FL substitutions (vowels and consonants) and low FL substitutions (vowels and consonants). We paid particular attention to these deviation types in our analysis being guided by previous research findings about the effect of high FL deviations on listeners' perceptions and learners' oral performances (e.g., Munro and Derwing, 2006; Kang and Moran, 2014).

The descriptive statistics for the segmental features are given in **Table 3**. The table summarizes the means, standard deviations, and 95% confidence intervals of high and low FL deviations before and after immersion for each of the three proficiency levels in the study. **Figure 1** provides an additional visual representation of the distributions of these features. Note that the plot represents the distribution of High and Low FL deviations for each proficiency level before and after the immersion. The boxes in the middle of the figure represent the interquartile range, the dark line in the middle of each box is the median, and the blue dot is the mean. The outliers are indicated by the gray circles outside of the overall range.

High and Low FL Deviations

As can be seen in **Figure 1** and **Table 3**, intermediate students improved by reducing the amount of high FL consonant and vowel deviations; in contrast, beginning and advanced learners demonstrated a higher amount of such deviations in their speech samples after the ESL immersion. It is noteworthy that the intermediate group had the largest amount of high FL substitutions of the three groups in the pre-test, while beginners displayed the highest number of substitutions in the post-test. The advanced learners had fewer high FL deviations in the pre-test than the intermediate group; however, their scores were almost equal in the post-test. Comparing both types of deviations for each proficiency group, the general patterns revealed differences in the distribution of high and low FL divergences in the pre- and post-test conditions. **Figure 1** shows that while beginners improved on the low FL substitutions, their production of high FL deviations increased. In contrast, intermediate learners demonstrated an opposite trend increasing low FL but reducing high FL divergences after immersion. The advanced group did not seem to show prominent changes in FL substitutions. None of the differences were significantly different as indicated by the 95% CIs overlapping by more than half in **Table 3** (see Cumming, 2009 for a detailed discussion). The only marginally significant result emerged for the beginner group in the amount of high FL deviations, $t = 1.97$, $p = 0.51$.

After fitting the linear mixed effects model on the data to examine high FL errors, the summary *F*-statistic obtained for the model was not significant $F(5,84) = 0.473$, $p = 0.79$. No significant interaction between the two fixed effects (Time and Proficiency) was observed, $F(2,42) = 1.69$, $p = 0.19$ indicating that the immersion did not necessarily improve the participants' pronunciation of FL-based segmental features across proficiency levels with regard to high

TABLE 3 | Summary of high and low functional load deviations for each proficiency group pre- and post-immersion.

	High FL deviations											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	3.37	2.26	2.23	4.51	4.57	3.72	2.69	6.45	3.55	1.97	2.55	4.55
Post-immersion	4.37	2.54	3.08	5.66	3.64	2.22	2.52	4.76	3.81	4.20	1.68	5.94

	Low FL deviations											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	5.48	4.08	3.42	7.54	4.21	2.75	2.82	5.60	4.80	2.86	3.35	6.25
Post-immersion	4.55	3.25	2.91	6.19	4.80	4.50	2.52	7.08	4.60	4.34	2.40	6.80

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

functional load divergences. The main effects of Proficiency and Time (pre/post) were not significant ($p = 0.83$), meaning that neither the high FL-based segmental changes happened significantly over 15 weeks, nor did proficiency make a difference. Overall, the fixed effects of Proficiency and Immersion, together with the random effects, explained close to 57% of the variance in High FL substitutions (conditional $R^2 = 0.569$). However, the fixed factors by themselves accounted for only 2% of the variance (marginal $R^2 = 0.022$) suggesting that the majority of variance was explained by participant factors and their L1s. To calculate the relative variance explained by the random factors, we transformed the variance values into proportions. The random idiosyncrasies of participants explained additional 42.5% of the variance; however, only 14% of variance was explained by the two L1s in the sample. No post hoc

analysis was performed because neither interaction effect nor main effect was significant in this model.

Another similar LMEM was computed to examine the effect of proficiency and immersion on the production of low functional load substitutions. The results summarized in **Table 3** above showed that the average number of low FL substitutions did not change noticeably for each group. Although the standard deviation in each group was quite large, on average, the beginner and advanced learners made fewer deviations of this nature in the post-immersion speech collection, and the intermediate group made low FL substitutions more frequently after the immersion. Similarly to high FL substitutions, the model did not reveal a significant interaction between the two fixed factors of proficiency and immersion, $F(2,42) = 0.46, p = 0.63$.

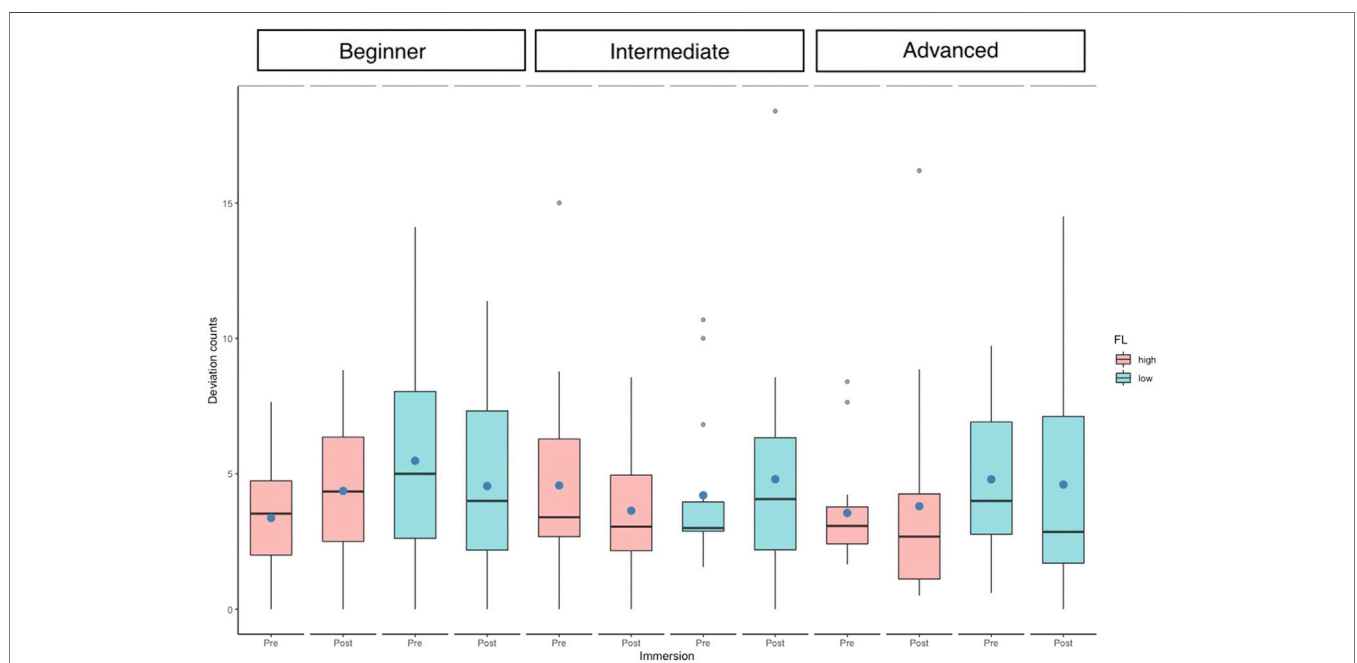


FIGURE 1 | Summary of results for segmental deviations per proficiency group before and after immersion.

TABLE 4 | Summary of speaking rate for each proficiency group pre- and post-immersion.

	Speaking rate (syllables per second)											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	2.16	0.50	1.96	2.36	2.35	0.53	2.14	2.56	2.50	0.37	2.35	2.65
Post-immersion	1.96	0.36	1.82	2.10	2.48	0.45	2.30	4.66	2.81	0.35	2.67	2.95

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

TABLE 5 | Summary of silent and filled pauses across proficiency levels pre- and post-immersion.

	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
	Silent pauses											
Pre-immersion	0.39	0.10	0.35	0.43	0.51	0.11	0.47	0.55	0.45	0.09	0.42	0.49
Post-immersion	0.46	0.09	0.43	0.49	0.49	0.08	0.46	0.52	0.47	0.12	0.42	0.52
	Filled pauses											
Pre-immersion	0.24	0.13	0.19	0.29	0.19	0.09	0.16	0.23	0.20	0.11	0.16	0.24
Post-immersion	0.15	0.10	0.11	0.19	0.24	0.11	0.20	0.28	0.26	0.15	0.20	0.32

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

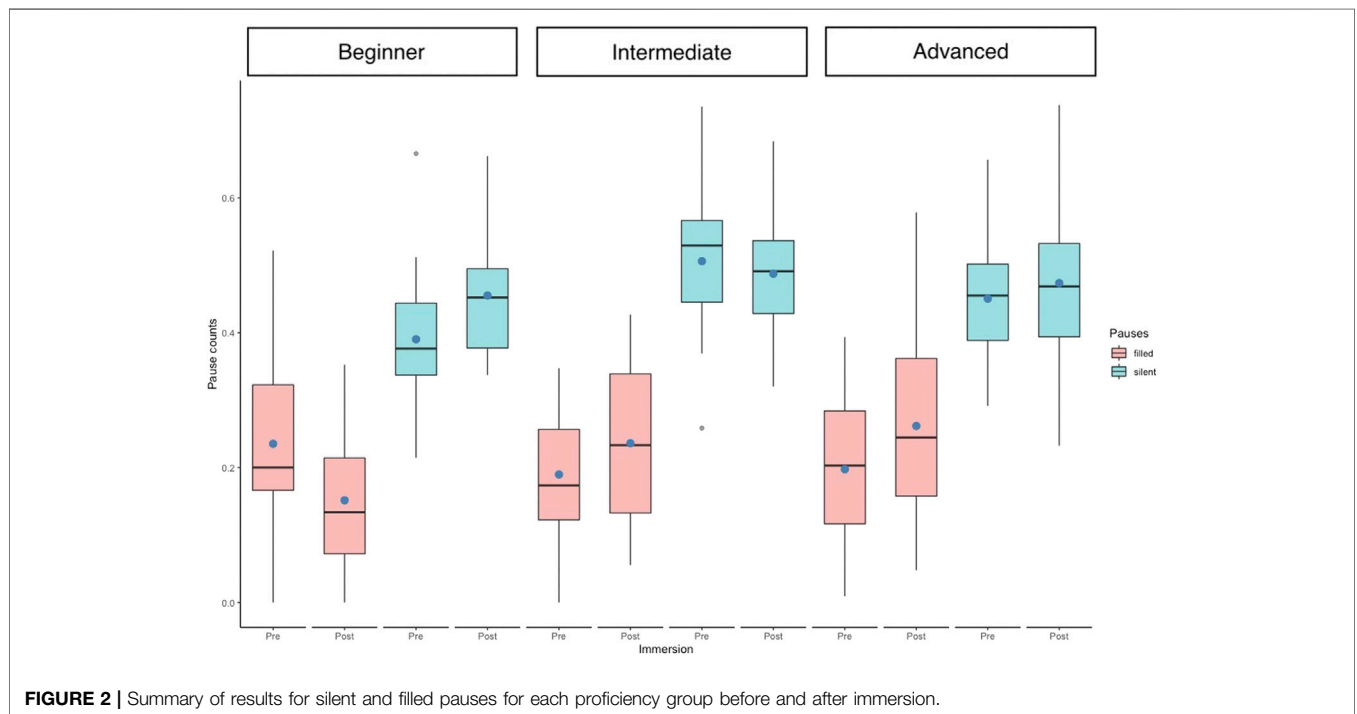


FIGURE 2 | Summary of results for silent and filled pauses for each proficiency group before and after immersion.

No significant main effects of time (pre/post), $F(1,42) = 0.8834$, $p = 0.35$, or Proficiency, $F(2,42) = 0.10$, $p = 0.90$ were detected either. Overall, this LMEM accounted for 72% of variance in the dependent variable (conditional $R^2 = 0.718$), although the fixed effects are responsible for only 1% of this variance (marginal $R^2 = 0.009$). The random participant factors explained over 70% of the variance. There was a small percentage of variance (3.2%) accounted for by participants' L1.

As stated earlier, additional LMEM analysis was done separately for high FL vowels and consonants as well as for low FL vowels and consonants. The emerged trends were similar to the main analysis presented here with none of the models being significant ($F(5,84) = 1.574$, $p = 0.176$ for high FL vowel deviations, $F(5,84) = 0.229$, $p = 0.949$ for low FL vowel deviations, $F(5,84) = 0.849$, $p = 0.519$ for high FL consonant deviations, and $F(5,84) = 1.265$, $p = 0.287$ for low FL consonant deviations).

Development of Suprasegmentals

The next five LMEM tested the effect of immersion on the development of suprasegmental features of pronunciation, namely, frequency measured by the speaking rate and the number of silent and filled pauses, stress measured by space, and intonation measured by pitch range. As stated earlier, the complete dataset of 150 speech files was used in building the five models. **Tables 4, 5** as well as **Figure 2** present the summaries of descriptive statistics for the three fluency measures.

Syllables per Second

The summary of descriptive statistics for the speaking rate in **Table 4** indicates that the intermediate and advanced learners showed improvement in the speaking rate in the post-immersion speech collection whereas the beginners were slower. There is also a clear difference between the proficiency levels and their respective average speaking rate with beginners being the slowest and the advanced learners being the fastest.

The LMEM with speaking rate as a dependent variable was significant, as indicated by the summary statistic, $F(5,150) = 12.47, p < 0.01$. Moreover, the interaction between proficiency and time of testing (pre- and post-immersion) was also significant, $F(2,75) = 8.54, p < 0.01$. In order to find out which proficiency groups significantly improved this aspect of fluency after immersion, post hoc pairwise comparisons using Tukey HSD were calculated. The results revealed that advanced learners were significantly faster after immersion than before, $t = 3.49, p < 0.01$. This result revealed Cohen's $d = 0.86$, which is considered a medium to large effect based on meta-analytically determined effect size guidelines for applied linguistics (Plonsky and Oswald, 2014). However, the speech rate of intermediate students did not change much showing a small effect size, $t = -1.53, p = 0.65$, Cohen's $d = 0.35$. The slow-down in beginners' speech was also not significant, $t = 2.16, p = 0.27$ with small to medium Cohen's $d = 0.50$. Proficiency and immersion, together with the random factors, were able to explain 60.5% of the variance (conditional $R^2 = 0.605$) and the fixed factors by themselves accounted for almost half of that variance (29%, marginal $R^2 = 0.286$). The participant differences explained most of the variance that occurred due to random factors while L1 background contributed less than 1% to the model.

Number of Silent and Filled Pauses

Table 5 below summarizes the average number of silent and filled pauses in the speech samples normalized per second of speech. In terms of the silent pauses in the speech samples, **Table 4** illustrates that only the intermediate group demonstrated a decline in their amount after ESL immersion. The participants from the other two levels used more silent pauses in the post-test, although the increase was barely noticeable. Interestingly, the number of silent pauses per second seemed to increase from beginner to intermediate speakers before the immersion and became less prominent after the immersion indicating that the distribution of silent pauses in the speech of beginner and

intermediate learners became more similar. In this study, a silence of over 0.1 s was considered a pause. It is possible that the intermediate and advanced learners produced more pauses, but they were much shorter than those of beginners. The analysis of filled pauses, also known as hesitation markers, in the speech samples across proficiency levels was inconclusive. While the beginner students used fewer filled pauses in their speech, both intermediate and advanced learners exhibited more filled pauses, as shown in **Table 5**.

Figure 2 below offers a visual comparison of the changes in the production of the two pause types by students from the three proficiency levels. It is noteworthy that silent pauses seem to be overall more frequent than filled pauses across all three groups in both pre- and post-tests.

The LMEM with silent pauses as a dependent variable was significant, $F(5,150) = 4.18, p < 0.01$. The model also uncovered a significant interaction between the fixed effects, $F(2,75) = 4.218, p = 0.018$. However, the examination of the results of post hoc Tukey HSD tests indicated that none of the groups displayed significant developmental changes that resulted from immersion, and the significant results in the fixed effects in the model were merely caused by the differences in the number of silent pauses between the levels in the pre- and post-tests. The model was able to explain a total of 61% (conditional $R^2 = 0.614$) of the variance in the number of silent pauses produced by the speakers with the fixed effects accounting for 8.5% of the total difference (marginal $R^2 = 0.085$). The results further indicated that both of the random factors explained almost 26% of the variance each.

The LMEM with filled pauses was statistically significant overall, $F(5,150) = 3.059, p < 0.01$, as well as the interaction between proficiency and the time of testing (pre/post), $F(2,75) = 9.305, p < 0.01$. The post hoc Tukey HSD tests indicated that only the beginner learners significantly improved their fluency by producing fewer filled pauses in the post-test, $t = 3.122, p = 0.029, d = 0.77$ (medium effect size). The model was able to explain 39% of the variance in the filled pauses across groups (conditional $R^2 = 0.39$) with the fixed factors of proficiency and immersion contributing to almost 9% of the variation (marginal $R^2 = 0.089$). The majority of the explained random variance in the data was accounted for by participant factors.

The results of the three LMEMs presented above focused on the fluency features of the speech samples, namely, syllables per second, number of silent pauses, and number of filled pauses. Based on our analysis, only advanced learners made substantial fluency progress, as indicated by significantly faster speech rate in the post-immersion test. Another significant change that was observed in the analysis was a decreased number of filled pauses in the speech samples produced by the beginner learners.

Space

In order to examine the change in prominence patterns in participants' speech after ESL immersion, another mixed effects model was built with space as the dependent variable. The boxplots in **Figure 3** show that all three groups used fewer prominent words in their speech in the post-test. In particular, the change of space in the intermediate group was particularly prominent followed by the advanced learners and beginners. The

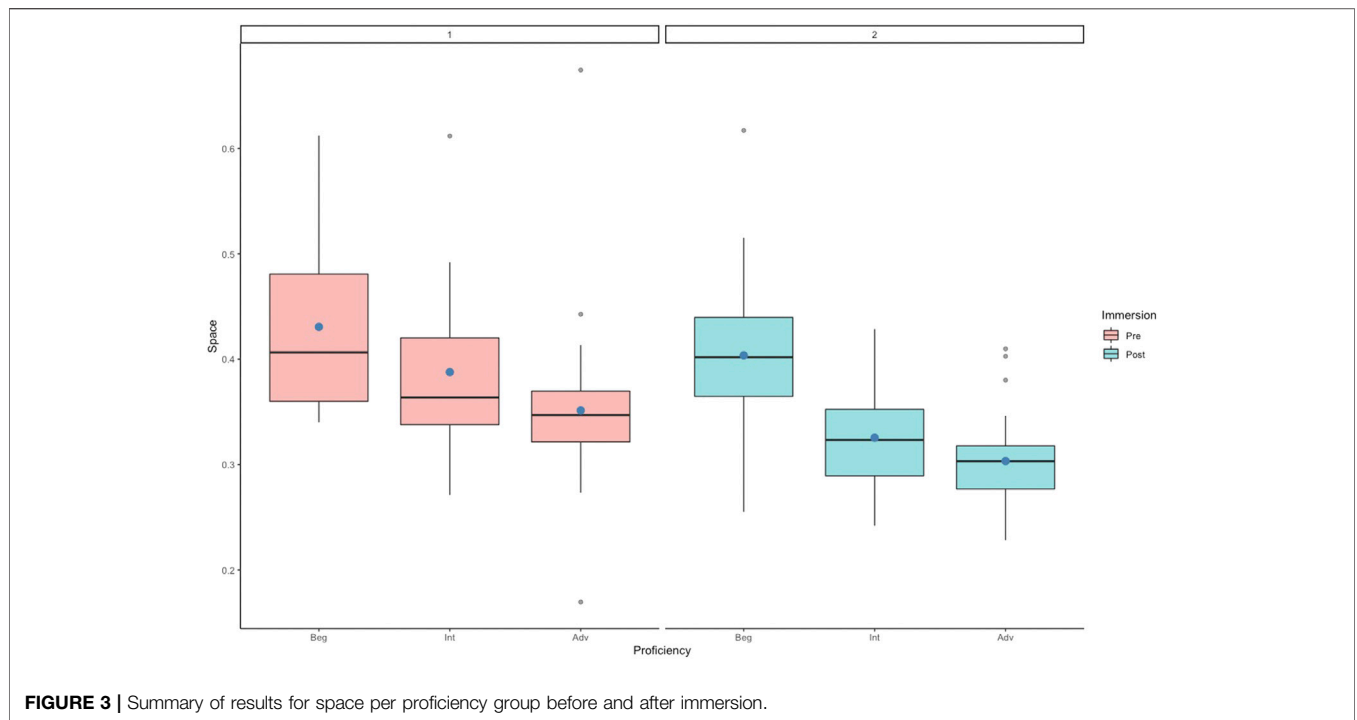


FIGURE 3 | Summary of results for space per proficiency group before and after immersion.

TABLE 6 | Summary of stress changes across proficiency levels pre- and post-immersion.

	Space											
	Beginner				Intermediate				Advanced			
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
		U	L			U	L			U	L	
Pre-immersion	0.43	0.08	0.40	0.46	0.39	0.07	0.36	0.42	0.35	0.08	0.32	0.38
Post-immersion	0.40	0.07	0.37	0.43	0.32	0.05	0.30	0.34	0.30	0.04	0.28	0.32

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

figure also reveals that the intermediate and advanced groups contained some outliers who stressed as few as 10% of words in a run or as many as 70% in the pre-test. However, the outliers were grouped closer to the average in the post-test (in the advanced group) or disappeared completely (in the intermediate group). The descriptive results for space given in Table 6 further exemplify that the stress changed across proficiency levels with more proficient learners stressing fewer words in a run and therefore improving their prominence.

Overall, the LMEM was significant, $F(5,150) = 12.86, p < 0.01$ as well as the interaction between the fixed effects, $F(2,75) = 17.26, p < 0.01$ suggesting that groups improved their prominence as a result of immersion. More specifically, according to the results of Tukey HSD tests, both intermediate and advanced groups significantly reduced their number of prominent words per run with a large and medium effect of immersion ($t = 3.858, p < 0.01, d = 1.15$ and $t = 3.148, p = 0.027, d = 0.79$ for intermediate and advanced groups, respectively). The change in space of the beginner group was not significant. The model accounted for nearly 52% of the variance (conditional $R^2 = 0.52$).

Over half of the variance was due to the fixed factors, as indicated by a large marginal effect size ($R^2 = 0.29$). Most of the random factor variance was explained by the participant factors.

Pitch Range

The last linear mixed effects model in the data analysis involved the speakers' pitch range as the dependent variable. Table 7 shows that the participant data in the three proficiency levels followed similar trends before and after the immersion with the advanced group demonstrating the widest range in contrast to the other two groups. It is noteworthy, however, that their pitch range was narrower in the post-test compared to the pre-test for this level. Similarly, in the post-test, the intermediate group also showed narrower while the beginners' pitch range was wider. The boxplots in Figure 4 additionally illustrate that some of the speakers displayed pitch range as wide as almost 250 Hz in the advanced group.

The fitted linear mixed effects model was significant according to the summary statistic, $F(5,150) = 2.324$. The results of the linear mixed effects model detected a significant effect of

TABLE 7 | Summary of pitch range changes across proficiency levels pre- and post-immersion.

	Pitch Range											
	Beginner			Intermediate				Advanced				
	M	SD	95% CI		M	SD	95% CI		M	SD	95% CI	
			U	L			U	L			U	L
Pre-immersion	74.7	38.6	59.6	89.8	82.2	39.9	66.6	97.8	104	50.4	84.2	124
Post-immersion	76.5	38	61.6	91.4	80.5	36.6	66.2	94.8	99	41.8	86.2	115

CI = confidence interval, U = upper 95% CI, L = lower 95% CI.

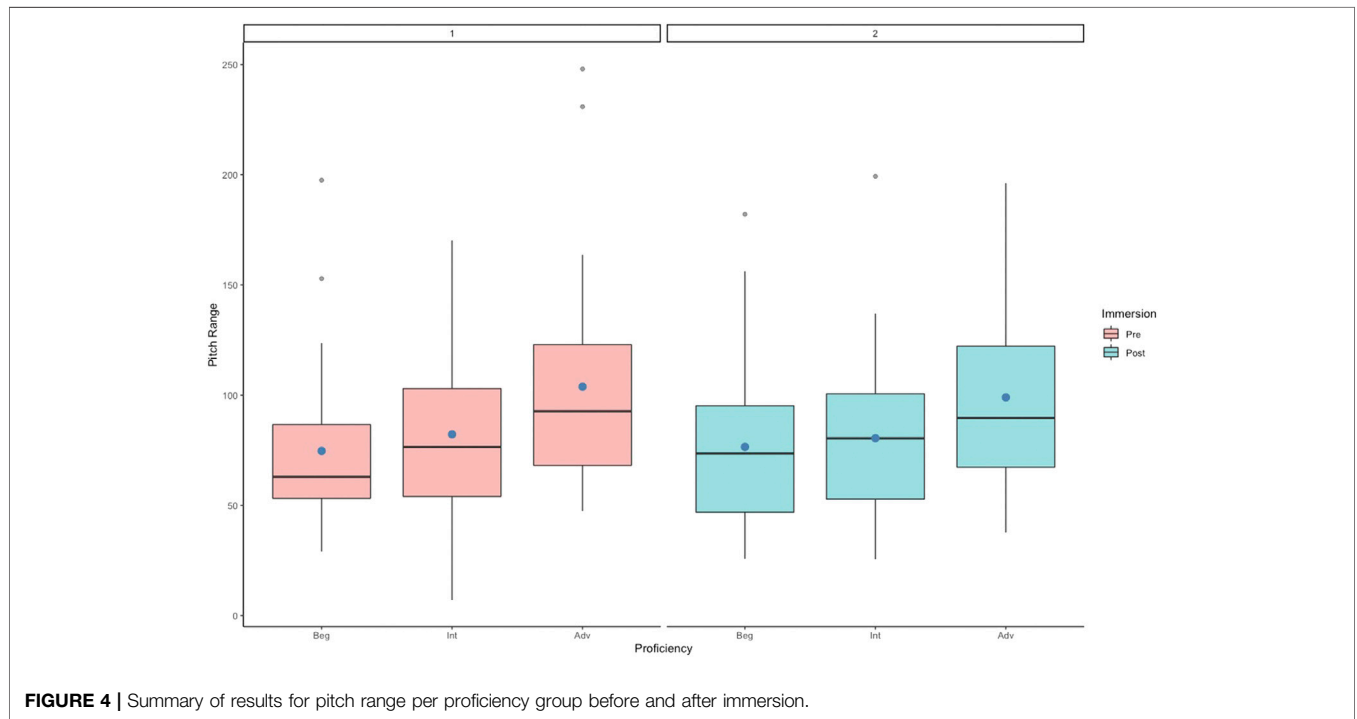


FIGURE 4 | Summary of results for pitch range per proficiency group before and after immersion.

proficiency $F(2,75) = 3.54, p = 0.03$ but not immersion $F(1,75) = 0.05, p = 0.82$ on pitch range. The interaction between the two effects was not significant. That is, the differences in the pitch range observed in the data occurred solely due to participants' proficiency and not as a result of ESL immersion. Overall, the model explained 61.5% (conditional $R^2 = 0.615$) of the variance in pitch range with the two fixed factors accounting for 7% (marginal $R^2 = 0.069$) of the variance and the remaining 54.5% falling mostly under the random effect of participants.

DISCUSSION AND CONCLUSION

The present study sought to offer a comprehensive investigation of the effects of a 15-week ESL immersion on learners' segmental and suprasegmental pronunciation features. Specifically, the study attempted to find out whether ESL immersion experience may play a role in reducing functional load divergences as well as making the participants' speech more fluent and intelligible. The results of the study revealed that

immersion had no significant effect on segmental deviations across proficiency levels. In fact, both immersion and proficiency were able to explain only 1–2% of the difference in learners' production of high and low FL substitutions over time. Previous studies (e.g., Kang and Moran, 2014) focused on the difference in FL vowel and consonant substitutions across proficiency levels and found significant differences. In the current study, however, the focus was not on the differences between the levels but on whether or not FL-based segmental features change over time at each proficiency level; nevertheless, significant changes did not emerge.

In terms of changes in pronunciation at the suprasegmental level, the overall findings suggest that there are clear improvements in some of the suprasegmental aspects of L2 speech. In particular, the speaking rate of the advanced group increased significantly over the period of 15 weeks. It is widely known that speaking rate plays a crucial role in perceptions of L2 speech comprehensibility and accentedness. Indeed, there exists a curvilinear relationship between speech rate and listener ratings with speech that is too slow or too fast being rated less

comprehensible and more accented (Munro and Derwing, 2001; Trofimovich and Baker, 2006). Although none of the groups in our study reached the optimal rate for accentedness and comprehensibility (4.76 and 4.23 syllables per second, respectively) suggested by Munro and Derwing (2001), the advanced learners made the most progress toward this goal. Another fluency feature that showed development in the study was filled pauses. In particular, beginner learners employed significantly fewer filled pauses in their speech after immersion. The finding supports recent research that showed that filled pauses can improve over time (e.g., Kang et al., 2021 in press). Furthermore, there was an improvement observed in intermediate and advanced learners' speech regarding their stress pattern (i.e., proportion of prominent words to the total number of words); that is, the number of prominent words significantly decreased over time for both of these groups. This decrease is especially beneficial for perceived accentedness of L2 speech. That is, the less a speaker stresses syllables in a sentence, the less accented they are found by the listeners (Kang, 2010). In contrast, the prosodic stress patterns of beginners did not improve. This finding lends potential support to the idea that the amount of L2 experience may influence the production of appropriate stress, as noted by Trofimovich and Baker (2006). It also is similar to Kang et al.'s, 2021 (in press) study where EFL learners improved their stress patterns most noticeably after 12 weeks of study in test preparation courses.

The two other suprasegmental features measured in our study (number of silent pauses and pitch range) did not exhibit substantial improvements after immersion. The overall pause structure of the participants' speech did not change with time for the three proficiency levels. The pitch range also stayed mostly the same across the groups. This result may indicate that the immersion experience does not affect all the suprasegmental features of speech in the same way and that some of these features, such as speaking rate and the distribution of silent pauses, may be more prone to improvement than others when students are immersed in an ESL environment. It is also possible that some suprasegmental features require more time and explicit instruction to develop noticeably (Levis, 2005). Taken together, the findings offer support to the idea that suprasegmental changes in L2 pronunciation may take less time than improvements in segmental features (Flege et al., 1997; Baker et al., 2001; Kang et al., 2021 in press).

An interesting finding that emerged as a result of the present study was the variance in the production of segmental and suprasegmental features that could be explained by participant factors in general and their L1 background. As a matter of fact, the possible individual differences among the participants accounted for the majority of the variance in the analyses, for example, over 50% in case of pitch range and 26% in the distribution of the silent pauses. The participants' L1 was not as pervasive of a predictor, although it did explain the other 26% of the variation in the use of silent pauses. These findings point at several inferences. First, the fact that the speakers' L1 explained so much of the variation in the silent pause distribution may be a sign of L1 transfer. Native speakers of Arabic and Mandarin may employ pauses differently in their first language and these patterns may be transferred from

the participants' native language to English (e.g., Ortega-Llebaria and Colantoni, 2014). More importantly, however, the findings reinforce the influence of learners' individual differences in the acquisition of L2 speech. Research has repeatedly shown that both external factors such as age of acquisition and L1–L2 distance as well as the internal factors of motivation, attitude, and metalinguistic awareness are strongly connected to the learners' success in acquiring L2 pronunciation skills (Baker and Trofimovich, 2006; Derwing and Munro, 2013; Saito et al., 2020). Moreover, other community-based factors could potentially play a role in the development of the speakers' L2 pronunciation (Derwing and Munro, 2013). While the current findings cannot account for the specific outside activities and opportunities with the out-of-classroom community that might have affected the learners' pronunciation improvement, they do provide indirect support to the role of such activities for pronunciation development during immersion.

Taken together, the results in the present study yield evidence to three points related to ESL instruction, particularly in the context of immersion. First, explicit instruction is needed to improve L2 learners' pronunciation, especially on the segmental level. The participants in the study were students in an Intensive English Program that did not include a pronunciation class in its curriculum. The learners were enrolled in a listening and speaking class, which did not include targeted pronunciation activities beyond what was offered in the textbook. It has been previously shown that only in combination explicit pronunciation instruction can immersion be beneficial for learners (e.g., Lord, 2010). Moreover, the lack of a separate pronunciation class presents a challenge for learners' pronunciation development, especially since L2 textbooks are inconsistent in covering pronunciation (e.g., Derwing et al., 2012). It is not surprising that the only significant improvements demonstrated by the learners in this study were in fluency and sentence stress since L2 textbooks often weigh heavier toward suprasegmentals.

Second, comparing our results to previous research on ESL immersion, it appears that the duration of the immersion is another factor determining its effectiveness. It may be the case that the 15 weeks that the learners spent in the ESL environment in the present study was not enough to result in significant pronunciation improvement. For example, Trofimovich and Baker (2006) found that learners who spent 3 months abroad were significantly less fluent than those who were abroad for 3 years or more. Specifically, the authors found that the speakers' stress timing was related to the amount of L2 experience. This implies that the longer learners spend in the immersion context, the more prominent their pronunciation development is. On the other hand, Trofimovich and Baker also observed that the learners' production of L2 suprasegmentals was strongly correlated with their perceived accentedness no matter the duration of immersion. Since suprasegmentals present a learning challenge for the learners despite the length of their immersion experience, the need for explicit pronunciation instruction is reiterated.

Finally, the results presented here provide additional evidence to the developmental threshold hypothesis (Collentine, 2009). Each level in our study significantly improved their

pronunciation by the end of their ESL immersion at least on some of the suprasegmental features, although this improvement was more noticeable among intermediate and advanced learners, as evidenced by the larger Cohen's *d* effect sizes. Importantly, the learners in the beginner group in our study were not "true" beginners. They were able to read the speaking prompt in the pre- and post-test and produced short but cohesive monologues in response to it. Therefore, it seems that the results of this study reinforce the idea that immersing students into the target language environment starts being effective only when the L2 learners reach at least basic literacy (Segalowitz and Freed, 2004) and that higher-level learners (intermediate and above) might improve their pronunciation more readily.

Despite an attempt in the present study to provide a comprehensive picture of pronunciation enhancement in the ESL immersion context, there are many questions that are left unanswered. For example, including additional segmental and/or suprasegmental features, such as tone choices, into the analysis could give a more fine-grained representation of specific changes in L2 English pronunciation. Another potential avenue of research could include the examination of interactive discourse (e.g., dialogues) in contrast to the monologues that were used in this study. It is oftentimes the case that suprasegmental features of conversations are different from monologic speech; thus, an investigation of changes in suprasegmentals employed by L2 learners in interactive discourse may shed additional light on prosodic development of L2 speech. Additionally, some individual differences can be further investigated through a qualitative approach as some of the speech properties (e.g., fluency features) can be idiosyncratic patterns instead of L2 proficiency or learning progression. In this research, we did not focus on the differences in segmental and suprasegmental deviations across learners' L1s. A potential qualitative study could investigate the effects of learners' L1 background on the types of

pronunciation deviations produced by the learners. Finally, a word of caution needs to be included with regard to the sampling procedures in the study. The number of speech samples for analyses was larger in case of suprasegmental analyses compared to segmental analysis. While linear mixed effects modeling performs equally well with smaller samples, we cannot exclude the slight possibility that the observed differences in suprasegmental features could have occurred due to the larger sample size. Moreover, the participants in the present study were recruited through convenience sampling. Although this type of sampling is typical for immersion studies, it would be beneficial for the domain if future research engaged in exploring pronunciation development in ESL immersion through random sampling.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northern Arizona University IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Audacity Team (2020). *Audacity(R): free audio editor and recorder [computer application]*. Available at: <https://audacityteam.org/> (Accessed September 10, 2020).
- Avello, P. (2010). "The effect of study abroad on Catalan/Spanish bilinguals' production of English /i-ɪ/&/æ-ʌ/contrasting pairs," in 6th International conference on language acquisition-CIAL, September 8–10, 2010 Spain, Europe: Universitat de Barcelona.
- Baker, W., Trofimovich, P., Mack, M., and Flege, J. E. (2001). "The effect of perceived phonetic similarity on non-native sound learning by children and adults," in Proceedings of the annual Boston University conference in language development. Editors B. Skarabela, S. Fish, and A. H-J. Do (Somerville, MA: Cascadilla Press), 36–47.
- Baker, W., and Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: the role of individual differences. *Int. Rev. Appl. Linguistics Lang. Teach.* 44 (3), 231–250. doi:10.1515/iral.2006.010
- Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Available at: <http://www.praat.org/> (Accessed October 13, 2020).
- Brecht, R., Davidson, D., and Ginsberg, R. (1995). "Predictors of foreign language gain during study abroad," in *Second Language Acquisition in a Study Abroad Context*. Editor B. Freed (Amsterdam, Netherlands: John Benjamins), 37–66.
- Brown, A. (1991). "Functional load and the teaching of pronunciation," in *Teaching English Pronunciation: a Book of Readings*. Editor A. Brown (Abingdon, United Kingdom: Routledge), 221–224.
- Catford, J. C. (1987). "Phonetics and the teaching of pronunciation: a systemic description of English phonology," in *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Editor J. Morley (Alexandria, VI: TESOL), 87–100.
- Collentine, J. G. (2009). "Study abroad research: findings, implications, and future directions," in *The handbook of Language Teaching*. Editors M. H. Long and C. J. Catherine (Hoboken, NJ: Wiley), 218–233.
- Cumming, G. (2009). Inference by eye: reading the overlap of independent confidence intervals. *Stat. Med.* 28 (2), 205–220. doi:10.1002/sim.3471
- Davidson, D. E. (2010). Study abroad: when, how long, and with what results? New data from the Russian front. *Foreign Lang. Ann.* 43 (1), 6–26. doi:10.1111/j.1944-9720.2010.01057.x
- Derwing, T. M., Diepenbroek, L. G., and Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation?. *TESL Can. J.* 30, 22–44. doi:10.18806/tesl.v30i1.1124
- Derwing, T. M., and Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherlands: John Benjamins.
- Derwing, T. M., and Munro, M. J. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Q.* 39 (3), 379–397. doi:10.2307/3588486
- Derwing, T. M., and Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: a 7-year study. *Lang. Learn.* 63 (2), 163–185. doi:10.1111/lang.12000
- Derwing, T. M., Munro, M. J., and Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Appl. Linguistics* 29 (3), 359–380.

- Derwing, T. M., Munro, M. J., Thomson, R. I., and Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Stud. Second Lang. Acquis* 31 (4), 533–557. doi:10.1017/s0272263109990015
- Derwing, T. M., and Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Appl. Lang. Learn.* 13 (1), 1–17.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., and Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Lang. Learn.* 54 (4), 655–679.
- Derwing, T. M., Thomson, R. I., and Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System* 34 (2), 183–193. doi:10.1016/j.system.2006.01.005
- Fayer, J. M., and Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Lang. Learn.* 37 (3), 313–326. doi:10.1111/j.1467-1770.1987.tb00573.x
- Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Q* 39 (3), 399–423. doi:10.2307/3588487
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonetics* 25, 437–470. doi:10.1006/jpho.1997.0052
- Freed, B. F., Segalowitz, N., and Dewey, D. P. (2004). Context of learning and second language fluency in French: comparing regular classroom, study abroad, and intensive domestic immersion programs. *Stud. Second Lang. Acquis.* 26 (2), 275–301. doi:10.1017/s0272263104262064
- Højen, A. (2019). "Improvement in young adults' second-language pronunciation after short-term immersion," in *A Sound Approach to Language Matters*. Editors Jespersen, A. B., Hejná, M., and Sørensen, M. H. AU Library Scholarly Publishing Services, 543–560.
- Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Can. Mod. Lang. Rev.* 64 (4), 555–580. doi:10.3138/cmlr.64.4.555
- Isaacs, T., and Trofimovich, P. (2012). Deconstructing comprehensibility. *Stud. Second Lang. Acquis* 34 (3), 475–505. doi:10.1017/s0272263112000150
- Iwashita, N., Brown, A., McNamara, T., and O'Hagan, S. (2008). Assessed levels of second languagespeaking proficiency: how distinct? *Appl. Linguistics* 29, 24–49.
- Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL: Int. Rev. Appl. Linguist. Lang. Teach.* 28 (2), 99.
- Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Lang. Assessm. Quart.* 9 (3), 249–269.
- Kang, O., Ahn, H., Yaw, K., and Chung, S.-Y. (2021). Investigation of relationship among learner background, linguistic progression, and score gain on IELTS. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia. Available at: <https://www.ielts.org/teaching-and-research/research-reports>.
- Kang, O., and Ghanem, R. (2016). Learners' self-perception of target language study in overseas immersion. *Jlitr* 7 (5), 819–828. doi:10.17507/jlitr.0705.01
- Kang, O., and Moran, M. (2014). Functional loads of pronunciation features in nonnative speakers' oral assessment. *Tesol Q* 48 (1), 176–187. doi:10.1002/tesq.152
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38, 301–315.
- Kang, O., Rubin, D., and Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Mod. Lang. J.* 94 (4), 554–566. doi:10.1111/j.1540-4781.2010.01091.x
- Kang, O., Thomson, R. I., and Moran, M. (2020). Which features of accent affect understanding? exploring the intelligibility threshold of diverse accent varieties. *Appl. Linguistics* 41 (4), 453–480. doi:10.1093/applin/amy053
- Lafford, B. (2004). The effect of context of learning on the use of communication strategies by learners of Spanish as a second language. *Stud. Second Lang. Acquis.* 26, 201–226.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Q* 39 (3), 369–377. doi:10.2307/3588485
- Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Lang. Ann.* 43 (3), 488–503. doi:10.1111/j.1944-9720.2010.01094.x
- MacIntyre, P. D. (2007). Willingness to communicate in the second language: understanding the decision to speak as a volitional process. *Mod. Lang. J.* 91 (4), 564–576. doi:10.1111/j.1540-4781.2007.00623.x
- Martinsen, Rob A. (2008). Short-term study abroad: predicting changes in oral skills. *Foreign Lang. Ann.* 43 (3), 504–530.
- Muñoz, C., and Llanes, À. (2014). Study abroad and changes in degree of foreign accent in children and adults. *Mod. Lang. J.* 98 (1), 432–449. doi:10.1111/j.1540-4781.2014.12059.x
- Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Stud. Second Lang. Acquis.* 23 (4), 451–468. doi:10.1017/s0272263101004016
- Munro, M. J., and Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: an exploratory study. *System* 34 (4), 520–531. doi:10.1016/j.system.2006.09.004
- Ortega-Llebaria, M., and Colantoni, L. (2014). L2 English intonation. *Stud. Second Lang. Acquis.* 36 (2), 331–353. doi:10.1017/s0272263114000011
- Pérez-Vidal, C., Juan-Garau, M., and Mora, J. C. (2011). "The effects of formal instruction and study abroad contexts on foreign language development: the SALA project," in *Implicit and Explicit Language Learning: Conditions, Processes and Knowledge in SLA and Bilingualism*. Editors R. Leow and C. Sanz (Washington, DC: Georgetown University Press), 115–138.
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Q* 35, 233–255. doi:10.2307/3587647
- Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English Specif. Purp.* 23 (1), 19–43. doi:10.1016/s0889-4906(03)00020-6
- Plonsky, L., and Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Lang. Learn.* 64 (4), 878–912. doi:10.1111/lang.12079
- R Core Team (2020). *R: a Language and environment for statistical computing*. Vienna, Austria. Available at: <https://www.R-project.org/> (Accessed August 25, 2020).
- Saito, K., Macmillan, K., Mai, T., Suzukida, Y., Sun, H., Magne, V., et al. (2020). Developing, analyzing and sharing multivariate datasets: individual differences in L2 learning revisited. *Ann. Rev. Appl. Linguist.* 40, 9–25. doi:10.1017/s0267190520000045
- Segalowitz, N., and Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition. *Stud. Second Lang. Acquis.* 26, 173–199. doi:10.1017/s0272263104262027
- Stevens, J. J. (2002). The acquisition of L2 Spanish pronunciation in a study abroad context. [Unpublished Doctoral Dissertation]. Los Angeles (CA): University of Southern California.
- Suzukida, Y., and Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Lang. Teach. Res.* doi:10.1177/1362168819858246
- Tavakoli, P., and Skehan, P. (2005). *9. Planning and task performance in a second language*. Amsterdam, Netherlands: John Benjamins.
- Towell, R. (2002). Relative degrees of fluency: a comparative case study of advanced learners of French. *Int. Rev. Appl. Linguist.* 40 (2), 117–150.
- Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Second Lang. Acquis.* 28 (1), 1–30. doi:10.1017/s0272263106060013
- Wennerstrom, A. (2000). "The role of intonation in second language fluency," in *Perspectives on Fluency*. Editor H. Riggensbach (Ann Arbor, MI: University of Michigan Press), 102–127.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kostromitina and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.