Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

The contribution of speech timing, f0 change, and voice quality to perceived prosodic proficiency in L2: a cross-lingual perspective

Heini Kallio

Tampere University
heini.h.kallio@tuni.fi

Abstract

Various studies have remarked the difficulty of the English stress and intonation systems to language learners. The current study takes a holistic approach to EFL prosody by investigating the contribution of speech timing, f0 change, and voice quality to the prediction of prosodic proficiency in EFL speakers with four different yet typologically close L1s. This study continues from an earlier one that suggests syllable duration to be the best feature for capturing relevant prominence characteristics of EFL speakers. However, the results varied depending on speaker L1.

Acoustic parameters were analyzed from 216 read utterances produced by EFL speakers with either Czech, Slovak, Hungarian, or Polish as their L1. Generalized linear mixed models were used to inspect the effect of acoustic parameters and speaker L1 on prosodic proficiency assessed by 40 trained raters. The results show a significant effect of timing parameters on assessed EFL proficiency, but the effect is considerably stronger for Hungarian EFL learners than for the other groups. Parameters of f0 change, in turn, proved more significant for Hungarian and Polish than for Czech and Slovak EFL learners. Further research is suggested to scrutinize the L1 effect in producing EFL prosody.

Index Terms: EFL prosody, speech timing, f0 change, L2 proficiency

1. Introduction

Prosodic systems of languages differ substantially, and language learners may experience a transfer effect from their native language (L1) to the target language [1]. Studies have found learners of English to face difficulties in various aspects of prosody, including intonation and stress production [2, 3, 4, 5, 6]. On one hand prosodic errors tend to accumulate in non-native (L2) speech so that disfluencies affect stress production [7], on the other hand it is possible that inappropriate production of L2 stress patterns affect overall prosody. The purpose of the current study is to investigate the role of global prosodic features in perceived prosodic proficiency in English as a foreign language (EFL) with respect to speaker L1 and provide suggestions for further research about L1 effect on EFL prosody. The focus is on parameters of speech timing, f0 change, and voice quality. Previous comparative research on EFL prosody has usually included L2 learners from two typologically distant languages [8], focused on specific features such as stress placement [9], or observed the effect of speaker's L1 from the perspective of perceived foreign accent [10]. The current study investigates EFL speakers from languages that are typologically relatively close: Czech, Slovak, Hungarian, and Polish.

One fundamental aspect that affects prosody is the production of language-specific stress patterns: alternation of stressed and unstressed syllables form the basis for speech rhythm. English has varying word stress, which makes the stress contrasts relatively strong and stress patterns less predictable than in languages with fixed stress [2]. As for acoustic realizations of stress, the primary prosodic feature signalling prominence in English is said to be duration: stressed syllables in English are considerably longer than unstressed syllables [11]. As opposed to English, all the L1s in this study have fixed word stress. In Czech, Slovak, and Hungarian lexical stress falls on the initial syllable, in Polish on the penultimate one. In addition to the placement of stress, the languages in question vary in the strength of stress contrasts and how prosodic signals are used in prominence production. English has relatively strong stress contrasts compared to the four L1s in this study [2, 12, 13, 14]. While duration is an important stress cue in English, Hungarian, Czech, and Slovak have a phonemic quantity distinction for vowels (Hungarian also has such a distinction for consonants), which may weaken the role of duration as a signal for prominence in these languages [15]. For Hungarian, f0 is suggested to be the strongest and duration as the weakest indicator of lexical stress [13]. Similar results have been found for Czech and Slovak, although studies note the lack of clear prosodic marking of prominence for these languages [16, 17, 12, 18]. Polish, as opposed to the other L1s in this study, does not have a quantity distinction in its vowels. As for stress markers, Malisz and Wagner [19] found that f0 and intensity serve as main determinants of overall prominence in Polish, leaving duration less significant. However, shortening of unstressed vowels in Polish was found in [20].

There are relatively few studies on the production of English stress patterns by speakers of Czech, Slovak, Hungarian, and Polish. For Hungarian learners of English, studies have focused mainly on stress placement, explaining production errors with both fixed word stress and quantity-sensitivity [21, 22]. [23] found some evidence of L1 stress transfer in the speech of Polish learners of English, but a comprehensive study on the production of English stress by Hungarian and Polish speakers is still lacking. Czech and Slovak speakers of English have been found to face difficulties in using duration to differentiate stressed and unstressed syllables [24, 25].

In a recent study, [26] compared syllable-level stress realizations of native and L2 speakers of English using a continuous wavelet transform (CWT) based prominence estimation method. The results of [26] support the importance of duration as the primary cue in English stress production, but also indicate that speakers with different L1s may use different prosodic cues when producing stress in English: duration proved to be the best prominence estimation feature to distinguish Czech and Slovak EFL speakers by their proficiency, while f0 seems to be the most salient prominence feature for Polish speakers of English. For Hungarian speakers of English, in turn, intensity served as the most significant individual feature. Since the CWT method used in [26] captures both word and utterance level prominence but

10.21437/SpeechProsody.2024-103

does not reveal how individual prosodic features manifest in the speech signal, this study further investigates the role of these parameters in EFL speech proficiency with respect to speaker L1. The focus is on utterance-level parameters of speech timing, f0 change, and voice quality.

2. Methods

2.1. Speech data

The speech data consisted of 216 read utterances in EFL: 56 samples from Czech speakers (N = 14), 56 samples from Slovak speakers (N = 14), 48 samples from Hungarian speakers (N = 12), and 56 samples from Polish speakers (N = 14).

All participants read an identical narrative text in English. From each participant's recording, the same subset of four utterances listed in 1 differing in length, vocabulary, and structure were chosen for human assessments and acoustic analysis. The speech data and collection procedure is described in more detail in [26]. All utterance-sized speech samples were segmented automatically with WebMAUS [27] at the syllable level following the sonority sequencing principle [28] and the segmentations were checked manually.

Table 1: *Target utterances*

They said it was a five star hotel, but I wouldn't give it one star.
There were big ships travelling past, and the sea was all polluted and brown – it looked horrible.
It was a sort of greeny black colour, and as we looked more closely we realised that it was full of frogs.
Instead we were very surprised to see lots of different types of vegetables: carrots, peas, cabbage, and a big bowl of lettuce.

2.2. Assessments

Each utterance was assessed for prosodic proficiency by 19 to 20 raters (N=40) in a pseudo-randomized setting. Assessments were performed by university students majoring either in English or Phonetics. The raters' L1s were Czech (N=7), Slovak (N=7), Hungarian (N=13), and Polish (N=13), and most assessors estimated themselves as C1-level (proficient) English users. A six-level proficiency scale (ranging from A1 to C2) for phonological control from the CEFR descriptors [29] was used for grading the speech samples. Here we focus on the assessment of prosodic features, which is a subsection of CEFR scale for phonological control and pays attention to features such as word-level and phrasal stress, rhythm, and intonation.

Before providing their assessments, the assessors participated in an identical training session where they were familiarized with the rating scale and assessment protocol. The quality of the assessment data was tested by measuring the consistency between raters using Cronbach's alpha. The consistency between raters was considered high with the value of Cronbach's alpha .94. A more detailed description of the assessment protocol and data can be found in [26].

2.3. Acoustic parameters and statistical analysis

Acoustic parameters of timing, f0 change, and voice quality were derived with a Praat [30] script. Articulation rate (syllables/second) was used as a reference parameter, since several studies have proved speed of speech measures the most important indicators of L2 fluency or spoken proficiency (see, e.g., [31, 5, 26]).

Two timing parameters were computed from syllable durations: rate normalized pairwise variability index (nPVI) and normalized standard deviation of syllable duration (n∆S). Similar timing parameters have previously been applied to L2 speech research by [32, 33, 34, 35]. The n∆S was computed as a function of mean standard deviation of syllable duration within a speech sample. Similarly, nPVI was computed as a function of mean durational difference between consecutive syllables within a speech sample.

Two parameters of f0 change were selected for analysis: f0 range and mean f0 slope in semitones, both computed from 0.05-0.95 percentile using sample median f0 as reference. While locally bound phenomena such as pitch accents have received much attention, fewer studies have investigated global f0 changes in speech [36, 37, 38, 35].

In addition, harmonics-to-noise ratio (HNR) was selected as a general voice quality parameter. Instead of only vowel segments, HNR was measured from full utterances to characterize the degree of overall acoustic periodicity. This measure can therefore be affected by, e.g., non-modality among speakers, hesitations in the lower proficiency levels (possible hesitations, fillers and self-corrections were not discarded from acoustic analysis), or erroneous voicing of unvoiced phonemes.

The effect of acoustic parameters on prosodic proficiency ratings was studied using generalized linear mixed model with average ratings as dependent variables, z-scored acoustic parameters as predictors, and speaker L1 as a random variable. Multiple linear regression models were fitted separately for subsets based on speaker L1 in order to take into account the possible differences arising from the speakers' L1s. The simplest models were derived with a stepwise feature selection method using the stepAIC algorithm (implemented in the R package MASS, [39]) that compares the Akaike Information Criterion (AIC) of all possible models and selects the one with the least information loss.

Table 2: *Acoustic parameters and their operationalizations*

Parameter	**Operationalization**
nPVI	Normalized pairwise variability index: rate-normalized mean difference (ms) between consecutive syllables
n∆S	Rate-normalized mean standard deviation of syllable duration (ms)
ArtRate	Rate of syllables per second without pauses or other disfluencies
f0 range	Pitch range in semitones, computed from 0.05-0.95 percentile using sample median f0 as reference
f0 slope	f0 slope in semitones, computed from 0.05-0.95 percentile using sample median f0 as reference
HNR	Harmonics-to-noise ratio (dB), forward cross-correlation analysis

3. Results

Table 3 summarizes the results of the regression models with predictor t-values and respective significance levels based on p-values as well as the adjusted R^2 of the final models. The gen-

Table 3: *Summary of the linear regression models with predictor t-values and adjusted R^2s. p-values: 0.1–0.05', 0.05–0.01*, 0.01–0.001**, < 0.001***.*

Parameter	**Model with all L1s**	**CZ**	**SK**	**HU**	**PL**
nPVI	6.37***	-	3.10**	2.54*	2.97**
n∆S	-3.28**	-	-3.38**	-	-2.24*
ArtRate	4.56***	5.62***	2.57*	1.73'	-
f0 range	3.20**	1.78'	-	2.36*	2.53*
f0 slope	-	-	-	-	-2.84**
HNR	-	2.71**	-	-2.62*	-2.99**
Model R^2 (adj.)	0.34	0.36	0.57	0.45	0.28

eralized linear mixed model with speaker L1 as random effect yielded higher AIC value than the model without it, indicating that speaker L1 does not improve the model fit. However, models fitted separately for each L1 group yielded differences in parameter effects. The predictive power of the models also varied depending on L1.

In the model with all L1 groups, nPVI and articulation rate proved to be the most significant predictors of prosodic proficiency. The results indicate that higher nPVI value and articulation rate is related to higher prosodic proficiency. Of the f0 change parameters, f0 range contributed significantly to the prediction model. Interestingly, n∆S showed negative effects for proficiency ratings. The overall model accounted for 34 per cent of the variance in ratings.

For the Czech L1 group, articulation rate was the most significant predictor of prosodic proficiency (t-value = 5.62, $p < 0.001$). f0 range showed a somewhat significant positive effect (t-value = 1.78, $p < 0.1$). Interestingly, harmonics-to-noise ratio also showed a significant positive effect for this group (t-value = 2.71, $p < 0.01$). The model accounted for 36 per cent of the variation in ratings.

For the Slovak speakers, n∆S was the most significant parameter with a negative t-value of -3.38 and $p < 0.01$, indicating that the more variation in syllable durations the lower the proficiency. Articulation rate and nPVI showed significant positive effects for proficiency. The predictive power of the model was the strongest for the Slovak L1 group with adjusted R^2 of 0.57.

For the Polish speakers, f0 slope provided a significant negative effect (t-value = -2.84, $p < 0.01$) while f0 range yielded a significant positive effect (t-value = 2.53, $p < 0.05$). As with Slovak and Hungarian L1 groups, nPVI had a significant positive effect on Polish speakers as well (t-value = 2.97, $p < 0.01$). n∆S also contributed to the prediction model with a negative yet significant effect (t-value = -2.24, $p < 0.05$). The predictive power of the model was the lowest for Polish L1 group with adjusted R^2 of 0.28.

The effects of nPVI and f0 range on proficiency ratings are the most consistent across L1 groups and are therefore selected for closer investigation. Figure 1 shows the effect of nPVI and figure 2 the effect of f0 range to average proficiency ratings for all EFL speakers pooled together as well as for each L1 group separately. The linear relationship between ratings and nPVI is the stronger for Slovak and Hungarian speakers than for Polish speakers, while the parameter shows no effect for the Czech group. f0 range, in turn, shows positive tendencies for all other L1 groups but the Slovak EFL speakers.

4. Discussion and conclusions

The goal of the study was to investigate the contribution of speech timing, f0 change, and voice quality to the prediction of prosodic proficiency in EFL speakers with four different yet typologically close L1s. Previous results on L2 speech show that the significance of timing parameters in L2 speech is greater when the target language and the speaker L1 differ in rhythmic properties [32, 33, 34, 35]. Since all language learners in this study come from L1s that differ from English with regards to rhythm, timing parameters (nPVI and n∆S) were expected to have an effect on proficiency ratings. When pooling all L1 groups together, nPVI showed the strongest positive effect on ratings, indicating higher proficiency when L2 speaker produced higher duration contrasts between consecutive syllables. This finding is in line with previous studies on EFL [24, 25, 26]. When analyzing each L1 group separately, however, nPVI had a significant positive effect on ratings for Slovak, Hungarian, and Polish speakers but not for the Czech speakers, contradicting previous results of [25] and [26]. Since Czech and Slovak are close in terms of prosodic typology, it was assumed that the two L1 groups share similar parameter effects. The differences in significant parameters can stem from varying proficiency distribution between L1 groups: the proficiency of the participants was not controlled for prior data collection, and thus some L1 groups gained on average lower proficiency grades than others. For example, the majority of Czech speakers fall into B- and C-level proficiency categories, while the Slovak speech samples are skewed towards lower levels, as can be seen from figures 1 and 2. It is likely that differences in proficiency distributions affect the results for speech timing parameters, since lower level speakers may focus more on segment- or word-level pronunciation and avoiding reading errors, which may lead to inappropriate stress patterns and cause a staccato-like rhythm in speech. Thus, to answer the question whether EFL speakers differ in their nPVI values with respect to their L1, it would be justifiable to study only speakers within the same proficiency levels. It should also be noted that possible effect arising from phrase-final lengthening on nPVI values was not acknowledged in the present analysis.

The results also suggest that the significance of commonly used speed measures such as articulation rate on spoken L2 proficiency should be treated with caution: in the present study, nPVI proved to be more significant predictor than articulation rate among most EFL learner groups. This is an important observation for future studies aiming to generalize spoken EFL proficiency for, e.g., automatic assessment purposes.

As for f0 change parameters, f0 range showed a significant positive effect in the prediction model with all EFL learners except for the Slovak L1 group. The finding, however, supports

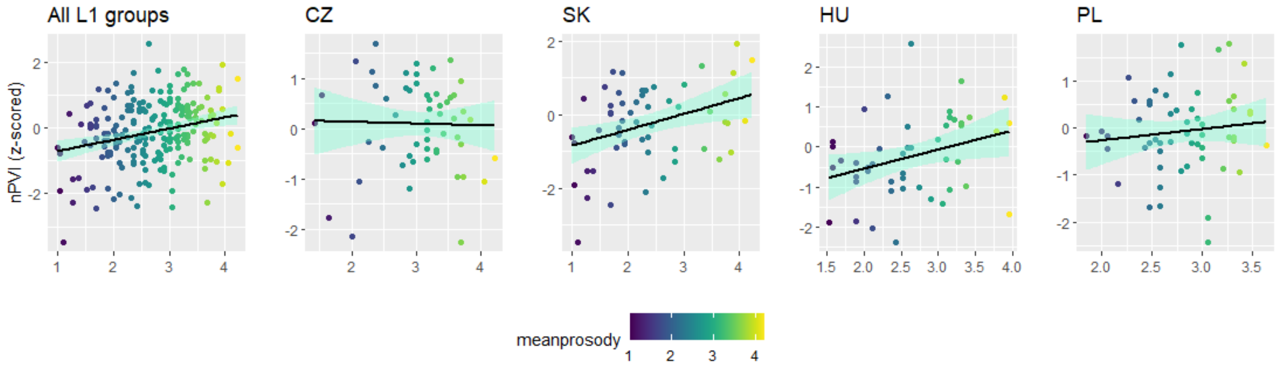

Figure 1: *nPVI values and mean prosodic proficiency ratings for all EFL speakers pooled together and for L1-specific speaker groups.*

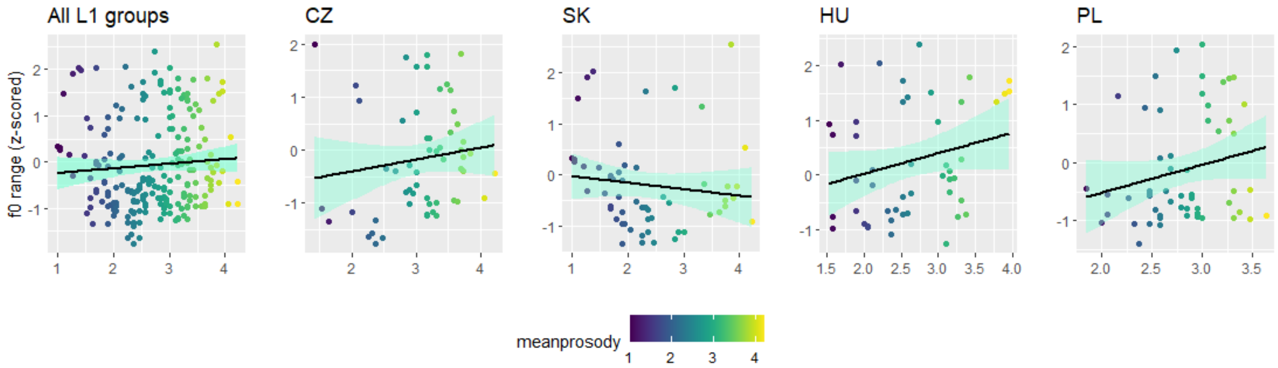

Figure 2: *f0 range and mean prosodic proficiency ratings for all EFL speakers pooled together and for L1-specific speaker groups.*

previous results on EFL speech [37]. In this study, the effect of f0 range was the strongest for the Polish and Hungarian L1 groups. Interestingly, f0 slope proved significant only for the Polish L1 group but yielded a negative effect, indicating that the steeper the f0 slope in an utterance, the lower the proficiency rating. This could stem from language-specific f0 patterns: [40] found that Polish speakers use wider f0 range than English speakers. However, in the current study f0 range and f0 slope provided conflicting results for the Polish L1 group. As the parameters selected for this study provide general utterance-level information on f0 change, a more detailed analysis on the different EFL f0 patterns is recommended. Since the utterances under analysis differ with respect to length and structure, it would be appropriate to take into account utterance-specific f0 patterning in the future.

Harmonics-to-noise ratio (HNR) contributed to the prediction of prosodic proficiency but showed opposite effects depending of the speakers' L1. Higher HNR significantly improved the perceived proficiency of Czech speakers of English, while the effect was significantly negative for the Hungarian and Polish L1 groups. As HNR presents the degree of overall periodicity, the results can stem from various phenomena: e.g., non-modality among speakers, hesitations in the lower proficiency levels (possible hesitations, fillers and self-corrections were not discarded from acoustic analysis), or voicing of unvoiced phonemes. In the future, the role of voice quality in perceived EFL proficiency should be scrutinized, e.g., in terms of non-modality with respect to utterance structure.

The coefficients of determination (R^2) in the L1-specific prediction models are somewhat in line with the results of [26]: in both studies, the proficiency of the Polish L1 group was the most difficult to predict with the selected parameters. In the current study, the prediction model for the Czech group also yielded a considerable lower R^2 than the models for Slovak and Hungarian L1 groups. The reasons for these differences can stem either from group-specific variation in the acoustic parameters or between-group variation in the proficiency distributions: it has been noted that the higher the proficiency of the L2 speakers, the more challenging it is to distinguish neighbouring levels of a CEFR scale using acoustic parameters [41, 26].

This study began to uncover some L1 effects on EFL prosody as well as contributions of these acoustic realizations to prosodic proficiency with speakers of Czech, Slovak, Hungarian, and Polish. The focus was on global parameters of speech timing, f0, and voice quality that can be related to the production of stress patterns. The results of the current study indicate that prosodic differences can be found in the speech of EFL speakers even from typologically close L1s. However, within-group variation is apparent and differences in speech timing, f0 change, and voice quality should be scrutinized by comparing features between speakers of the same EFL proficiency level. Future research will focus on comparing utterance-specific speech timing, f0 features, and voice quality as well as segment-level phenomena of EFL stress realizations between the respective L1 groups.

5. References

[1] D. Hirst and A. Di Cristo, *Intonation systems: a survey of twenty languages*. Cambridge University Press, 1998.

[2] M. Halle and S. J. Keyser, *English stress*. Harper and Row New York, 1971.

[3] A. Wennerstrom, "Intonation as cohesion in academic discourse: A study of Chinese speakers of English," *Studies in Second Language Acquisition*, vol. 20, no. 1, pp. 1–25, 1998.

[4] L. D. Hahn, "Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals," *TESOL Quarterly*, vol. 38, no. 2, pp. 201–223, 2004.

[5] J. Kormos and M. Dénes, "Exploring measures and perceptions of fluency in the speech of second language learners," *System*, vol. 32, no. 2, pp. 145–164, 2004.

[6] J. Field, "Intelligibility and the listener: The role of lexical stress," *TESOL Quarterly*, vol. 39, no. 3, pp. 399–423, 2005.

[7] L. Rasier and P. Hiligsmann, "Prosodic transfer from L1 to L2. Theoretical and methodological issues," *Nouveaux cahiers de linguistique française*, vol. 28, no. 2007, pp. 41–66, 2007.

[8] A. Li and B. Post, "L2 acquisition of prosodic properties of speech rhythm: Evidence from l1 mandarin and german learners of english," *Studies in Second Language Acquisition*, vol. 36, no. 2, pp. 223–255, 2014.

[9] H. Altmann, *The perception and production of second language stress: A cross-linguistic experimental study*. University of Delaware Newark, DE, USA, 2006.

[10] T. M. Derwing and M. J. Munro, "Accent, intelligibility, and comprehensibility: Evidence from four l1s," *Studies in second language acquisition*, pp. 1–16, 1997.

[11] P. Ladefoged and K. Johnson, *A course in phonetics*. Nelson Education, 2014.

[12] Š. Beňuš, U. D. Reichel, and K. Mády, "Modeling Accentual Phrase Intonation in Slovak and Hungarian," in *Complex Visibles Out There*, L. Veselovská and M. Janebová, Eds. Olomouc, Czech Republic: Palacký University, 2014, vol. 4, pp. 677–689.

[13] I. Vogel, A. Athanasopoulou, and N. Pincus, "Acoustic properties of prominence in Hungarian and the Functional Load Hypothesis," in *Approaches to Hungarian: Volume 14: Papers from the 2013 Piliscsaba Conference*, K. É. Kiss, B. Surányi, and É. Dékány, Eds. John Benjamins, 2015, vol. 14, pp. 267–292.

[14] A. Cwiek and P. Wagner, "The acoustic realization of prosodic prominence in polish: Word-level stress and phrase-level accent," *Proceedings of Speech Prosody 2018*, 2018.

[15] B. Remijsen, "Lexically contrastive stress accent and lexical tone in ma'ya," *Laboratory phonology*, vol. 7, pp. 585–614, 2002.

[16] T. Dubeda and J. Votrubec, "Acoustic analysis of Czech stress: Intonation, duration and intensity revisited," in *Proceedings of the Ninth European Conference on Speech Communication and Technology*, 2005.

[17] R. Skarnitzl and A. Eriksson, "The acoustics of word stress in Czech as a function of speaking style," in *Proceedings of INTERSPEECH 2017*, 2017, pp. 3221–3225.

[18] Š. Beňuš and K. Mády, "Effects of lexical stress and speech rate on the quantity and quality of Slovak vowels," in *Proceedings of Speech Prosody 2010*, 2010.

[19] Z. Malisz and P. Wagner, "Acoustic-phonetic realisation of Polish syllable prominence: a corpus study," in *Rhythm, melody and harmony in speech. Studies in honour of Wiktor Jassem.*, 2012, vol. 14.

[20] A. Rojczyk, "Quality and duration of unstressed vowels in Polish," *Lingua*, vol. 217, pp. 80–89, 2019.

[21] J. Archibald, "The learnability of English metrical parameters by adult Spanish speakers," *IRAL: International Review of Applied Linguistics in Language Teaching*, vol. 31, no. 2, p. 129, 1993.

[22] M. Bakti and J. Bóna, "Source language-related erroneous stress placement in the target language output of simultaneous interpreters," *Interpreting*, vol. 16, no. 1, pp. 34–48, 2014.

[23] J. Archibald, "The acquisition of second language phrasal stress: a pilot study," *Language Acquisition and Language Disorders*, vol. 16, pp. 263–290, 1997.

[24] M. Bilá and J. Zimmermann, "English rhythm and the Slovak speaker," in *Proceedings of the International Congress of Phonetic Sciences 1999*, vol. 547, 1999, p. 550.

[25] L. Weingartová, K. Poesová, and J. Volín, "Prominence contrasts in Czech English as a predictor of learner's proficiency," in *Proceedings of Speech Prosody 2014*, 2014, pp. 236–240.

[26] H. Kallio, A. Suni, and J. Šimko, "Fluency-related temporal features and syllable prominence as prosodic proficiency predictors for learners of english with different language backgrounds," *Language and Speech*, vol. 65, no. 3, pp. 571–597, 2022.

[27] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[28] S. Parker, "Sonority," *The Blackwell companion to phonology*, pp. 1–25, 2011.

[29] Council of Europe, "Common European framework of reference for languages: Learning, teaching, assessment. Companion Volume with New Descriptors," Council of Europe, Education Department, Tech. Rep., 2017.

[30] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer.[Computer program]. Version 6.0. 19," *Online: http://www. praat. org*, 2016.

[31] T. M. Derwing, M. J. Rossiter, M. J. Munro, and R. I. Thomson, "Second language fluency: Judgments on different tasks," *Language Learning*, vol. 54, no. 4, pp. 655–679, 2004.

[32] L. White and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35, no. 4, pp. 501–522, 2007.

[33] T.-Y. Jang, "Automatic assessment of non-native prosody using rhythm metrics: Focusing on korean speakers' english pronunciation," in *Proc. of the 2nd International Conference on East Asian Linguistics*, 2009.

[34] U. Gut, *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Peter Lang, 2009, vol. 9.

[35] H. Kallio, M. Kautonen, and M. Kuronen, "Prosody and fluency of finland swedish as a second language: Investigating global parameters for automated speaking assessment," *Speech Communication*, vol. 148, pp. 66–80, 2023.

[36] R. Ullakonoja, "Comparison of pitch range in finnish (l1) and russian (l2)," in *Proceedings of the 16th International Congress of Phonetic Sciences, 6-10 August 2007, Saarbrücken, Germany.* Universität des Saarlandes, Saarbrücken, 2007.

[37] M. G. Busà and M. Urbani, "A cross linguistic analysis of pitch range in english l1 and l2." in *ICPhS*, 2011, pp. 380–383.

[38] M. Kuronen and E. Tergujeff, "Second language prosody and its development: connection between different aspects," *The Language Learning Journal*, vol. 48, no. 6, pp. 685–699, 2020.

[39] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth, and M. B. Ripley, "Package 'mass'," *Cran r*, vol. 538, pp. 113–120, 2013.

[40] B. Andreeva, G. Demenko, B. Möbius, F. Zimmerer, J. Jügler, and M. Oleskowicz-Popiel, "Differences of pitch profiles in germanic and slavic languages," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[41] H. Kallio, A. Suni, J. Šimko, and M. Vainio, "Analyzing second language proficiency using wavelet-based prominence estimates," *Journal of Phonetics*, vol. 80, p. 100966, 2020.