2022 ACTA UNIVERSITATIS CAROLINAE PAG. 115–129
PHILOLOGICA 1 / PHONETICA PRAGENSIA

ONLINE GUIDED PRONUNCIATION PRACTICE HELPS ADULT EFL LEARNERS IMPROVE L2 PROSODY

ŠÁRKA ŠIMÁČKOVÁ, VÁCLAV JONÁŠ PODLIPSKÝ
Palacký University Olomouc

ABSTRACT

This study tests the efficacy of a pronunciation course in developing advanced EFL learners' expressive reading during a semester of online instruction. The course, designed for future English-language professionals, emphasises primacy of perception before production, the importance of noticing phonetic detail, expert and peer feedback, and context-situated tasks. The magnitude of pitch movements and reading tempo were assessed before and after the course for a trained group, who received the pronunciation practice, and a comparison group attending a course about the theory and research of foreign accents in English. Only the Trained group's expressive prosody improved: the learners slowed down their delivery and produced the utterances with a wider pitch range. The results suggest that adult foreign language learners can benefit from pronunciation training in a distance learning environment.

Key words: distance learning, English pronunciation, expressive prosody, pitch range, reading tempo

1. Introduction

1.1 Research background

Prosody is essential to organising connected speech and it plays an important role in effective communication (Hirschberg, 2002). The prosodic structure of an utterance reflects the speaker's organisation of thought (in prosodic phrasing and prominence), the degree of certainty with which they are speaking (in melodic patterns), as well as pragmatic meanings that the speaker may want to convey beyond the lexical meaning of spoken words (such as doubt, surprise or irony). To the listener, prosodic cues indicate what to pay attention to when processing speech, what to anticipate in the upcoming discourse, or when to take their turn in conversation.

Mastering the prosody of a second language (L2) is not an easy feat. L2 prosodic learning is affected by first-language (L1) transfer just like the learning of L2 segments (Mennen & de Leeuw, 2014) and prosodic features of L2 speech often serve as markers of foreignness to a native listener's ear (De Mareüil & Vieru-Dimulescu, 2006). Non-na-

https://doi.org/10.14712/24646830.2022.32

tive prosody may negatively impact the reception of an L2 speaker's message, affecting listeners' interest in what is being said, or reducing comprehension (Kang, 2008; Kang, Rubin, & Pickering 2010). Although research on L2 prosody training is relatively limited, existing studies have shown that foreign language prosody can be improved through training (see Lengeris, 2012 for a review). In the speech of Czech learners of English, the learner population trained in this study, prosodic features (F0 variation and articulatory rate) have been shown to predict accent ratings (Volín and Skarnitzl, 2010), although the narrow pitch range often taken to be typical of Czech-accented English cannot be attributed purely to L1 interference (Volín, Poesová & Weingartová, 2013).

The current study tests the ability of young adult non-immersion Czech EFL learners to change prosodic aspects of their English speech as a result of guided pronunciation practice. Our learners are students at Palacký University training to become English language professionals, and as such are highly proficient and active EFL users. They opted for a 13-week-long pronunciation course and thus can be regarded as motivated to improve their spoken English. Since 2 weeks into the course the first covid lockdown closed all classrooms, the training moved online. It relied primarily on the Moodle platform and on individual audio messages. In this study the efficacy of such online guided pronunciation practice is tested by considering the learners' ability to read with expression, what is sometimes called "prosodic reading" and involves expressive rhythmic and melodic patterns (Dowhower, 1991). Our goal is to determine whether guided online pronunciation practice can lead to improvements in adult EFL learners' ability to read with prosody.

1.2 The training

The pronunciation training within this course targets both segmental and suprasegmental aspects of English pronunciation of Czech speakers. It is grounded in four core assumptions. First, while recognizing the complexity of the relationship between perception and production, we assume that accurate perception is important for developing accurate production (Baese-Berk, 2019; Derwing & Munro, 2015). Consequently, irrespective of the pronunciation feature targeted, the training always includes listening to (multiple and varied) samples of native English speech. The listening tasks guide learners' attention to, and awareness of, specific phonetic features in authentic audio input. Second, it is assumed that building up speech production skills involves proceduralization, i.e. progression from controlled to automatized performance (Gatbonton & Segalowitz, 1988). In each session, initial speaking activities (e.g. imitation, shadowing, or chanting) give the learners opportunities to practise English segmental or prosodic features without overburdening their attentional resources. Subsequently, more demanding tasks (e.g. narrations, rehearsed dramatic dialogues, impromptu role-plays) are introduced. Third, we strive to situate the practice in meaningful contexts to help the learners transfer pronunciation gains to actual language use (Lightbown, 2008). Fourth, we regard feedback as a force that drives learning by helping learners notice the gap between their own pronunciation and that of a native speaker model, thus leading to faster learning gains (Saito & Lyster, 2012).

Although the course had been taught 4 times and received positive evaluations from students, this was the first attempt to empirically test its impact on the learners' output.

The fifth course edition in spring 2020 was organised as follows: the first week's meeting, during which the pre-test data were collected by a research assistant, was followed by 12 training sessions, with post-test data submitted by the learners via email a week after the last session. The training activities addressed multiple aspects of English pronunciation: the perception and production of segmental features typical for the Czech accent in English, as well as phrase- and sentence-level prosody. Learners practiced not only tempo and intonation, the two features of interest for the current paper, but also lexical stress, rhythm, prosodic structuring of discourse, and emotive prosody, amongst other things. Activities targeting tempo and intonation included listening (meaning- and form-focused, of longer and shorter passages, of varied speakers or focused on a single speaker, self- and peer-listening), reading aloud, repetition, imitation, shadowing of varied speech samples, transferring the speech prosody of an example recording to new texts, role-playing based on model recordings, drama rehearsals, rehearsed and impromptu monologue deliveries, recitation of rhythmical verse, and gesture-supported productions. Because of the covid lockdown, the regular 90-minute face-to-face training was replaced by online activities. The course was based firstly on weekly Moodle postings of practice materials, i.e. audio files and handouts, secondly on learners' submissions of self-recordings of all oral tasks, and finally on the instructor's and peers' feedback. Each learner received individualised written and audio feedback on their submissions from the teacher. Selected parts of the learners' recordings were posted on Moodle for peer feedback, which had the form of comprehensibility and accentedness ratings, ratings of the proximity of learners' productions to a native model, or written commentaries on specific pronunciation features. Each week's session was completed by the instructor's feedback on the learner group performance.

1.3 Reading prosody

Good reading prosody reflects readers' consideration of the communicative purpose and it facilitates listeners' comprehension of what is being read. It helps parsing, provides discourse information, directs listeners' attention, adds emphasis, conveys emotions, and offers implicit information. In literature on reading skills of children, reading prosody is sometimes viewed as a component of reading fluency (Kuhn et al., 2010), sometimes fluency and expressivity are separated (Cowie et al. 2002), which is what we also find useful: it is a common experience to encounter an L2 learner who reads aloud fluently but still lacks expressivity in their reading.

Our learners could also read fluently and accurately, as subjectively evaluated by the instructor and one research assistant. They did occasionally produce false starts, self-corrections, hesitations or pauses, as expectable for an adult L2 speaker not trained in reading aloud. They were accurate in the sense of automatically recognizing and instantly producing all the words in the text. On the other hand, it was evident that the learners' ability to read with expressive prosody was rather weak, despite some individual variation. Poor expressive reading is not necessarily only the effect of reading in an L2; adults reading in their L1 also vary in the ability to read with prosody. However, we do think that non-nativeness contributed to reduced expressivity of in most of the learners in this course. Consider the example sentence *"It can't be the milkman because he came this* *morning, and it can't be the boy from the grocer because this isn't the day he comes.*" Figure 1 shows F0 of a native reading and of a learner pre- and post-training readings. The learner's initial lack of phrasing and pitch dynamics, as compared to native production, is clearly evident.

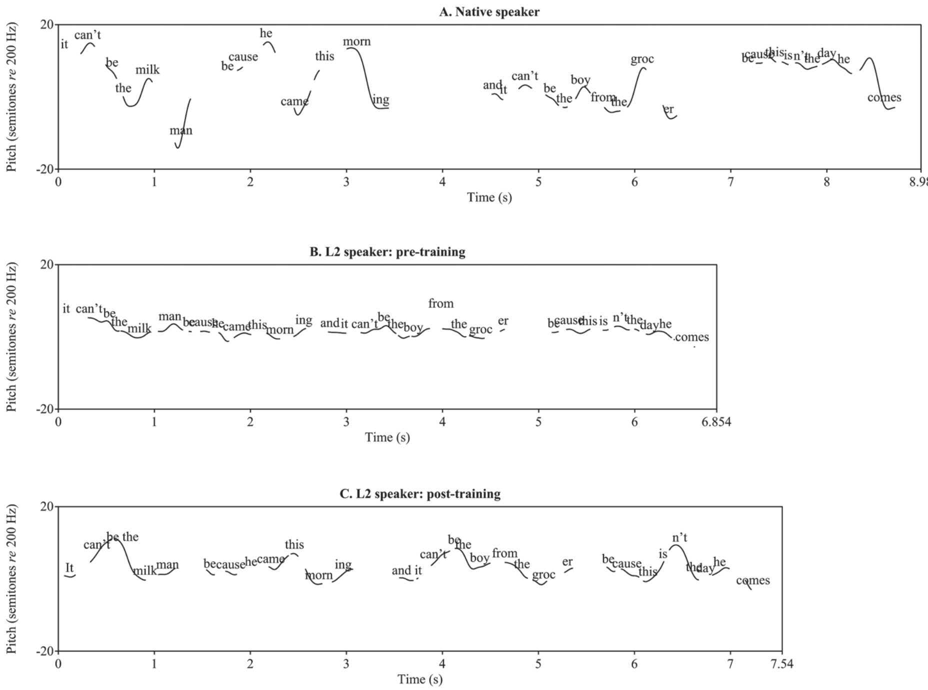

Figure 1. Example of F0 tracks from a native speaker and a learner.

1.4 Correlates of expressivity in this study

We focus on two prosodic correlates of expressivity that have been found relevant in research on the development of reading prosody in children (e.g. Cowie et al., 2002). One is F0 variation measured as the range from the 90th to the 10th percentile. In a subset of data, the magnitude of F0 movements in syllables bearing the nuclear pitch accent was also measured, in other words the difference between the F0 at the beginning of a fall or a rise and the F0 at the end. We expected an increase in expressivity in the post-test data that would be reflected in a wider pitch range.

The other correlate of expressivity considered here is reading tempo. We expected an increase in expressivity to be reflected in *slower* rather than faster reading. While slower speaking rate is often a marker of foreign accentedness (Munro & Derwing, 1998) and even advanced L2 learners are found to speak more slowly than native speakers (e.g. Huang & Gráf, 2020), native listeners may also judge non-native speech as too fast (Munro & Derwing, 2001). What is more, our study is not concerned with the overall speaking tempo but with a tempo that is appropriate for a specific task, namely for reading with expression. The instruction for the participants was to deliver a story as if for a young audience, in a way that should be engaging. We assume that fast reading tempo in this context actually reduces expressiveness and contributes to monotony. To illustrate the difference between native and non-native reading tempo, in Figure 1 we compare the duration of two renditions of an example target sentence: the native speaker's and an example learner's pre-training rendition. The learner's realisation is approximately 2.3s shorter, which is roughly the time it takes to produce 8-10 syllables. The difference is not only due to the more dramatic pauses evident in the native speaker's rendition. For example, the duration of the clause "*because he came this morning*," pronounced in the example as a fluent intonational phrase both by the learner and the native reader, is 1.43s and 1.74s respectively.

2. Methodology

2.1 Participants

Altogether 16 participants, aged 20 to 27 (mean 22), completed the pronunciation course. The 11 women and 5 men, all native speakers of Czech, were students majoring in English at Palacký University Olomouc. Their general proficiency level was relatively high, between C1 and C2 in CEFR (Council of Europe, 2001): 9 were undergraduate students and had passed the required C1 exam, the remaining 7 were graduate students preparing for the final C2 exam. These participants will be referred to as the Trained Group. Their performance was compared to that of the Comparison group of 14 participants, 9 women and 5 men drawn from the same learner population. They were 12 undergraduate and 2 graduate students, who attended the Foreign Accent seminar, a theoretical course dealing with linguistic, psycholinguistic and sociolinguistic aspects of foreign accents in English.

2.2 Recordings

The test consisted of reading the children's classic story "The tiger who came to tea" (Kerr, & McEwan, 2006). For the Trained Group, the pre-test took place during the first course meeting at Palacký University. The data were collected individually in a soundproof booth by a student assistant on a Zoom H4n recorder recording the speech at 16-bit and 44.1 kHz. Because of the Covid-19 restrictions, the post-test data were recorded by the learners on their mobile phones (M4A format) in quiet conditions and were generally of a relatively high quality. The Comparison Group data, collected in the spring 2021, were studio-recorded on both tests.

Both on the pre- and the post-test, the participants were instructed to read the story aloud in an engaging way to an imagined audience of pre-schoolers who would view it with pictures on YouTube. The learners were encouraged to rehearse reading the story from the actual picture book (Kerr & McEwan, 2006).

In addition to the Czech learners' data, we analysed the recordings from 7 native English speakers (5 women, 2 men) available on YouTube (see Online data resources). Six of the speakers were judged to be speakers of Southern British English (by a native speaker of that accent and by a Czech phonetician). One female speaker's pronunciation had features of a northern English accent. The analysis focused on the within-learner pre-vs-post-test improvements. The native speakers' data were used for baseline reference (see Figures 4 and 6).

2.3 Analysis

Altogether 16 intonational-phrase-size units, listed in Table 1, were analysed in terms of reading tempo and F0. Due to an error during the pre-test administration, only the first 12 phrases were available for the analysis in 2 participants' recordings in the Trained group. The complete dataset consisted of 944 phrases; 492 and 448 from the Trained and Comparison groups, respectively.

The phrases were all direct speech statements uttered in the story by 3 different characters – Mommy, Tiger, and Daddy. Utterances were coded as fluent or as involving a disfluency such as a perceptible hesitation or restarting a word. For each fluent utterance, the reading tempo was computed by dividing the number of spoken syllables in a phrase by the total time to read the phrase. The disfluent utterances were excluded from the analysis of tempo. In total 57 utterances were identified as disfluent (pre-test: 24 in the Trained and 13 in the Comparison group; post-test: 5 in the Trained and 15 in the Comparison group). Furthermore, F0 (in semitones *re* 200Hz) was estimated using Praat's autocorrelation algorithm, with all parameters set to default values except the 'Pitch floor' and 'Pitch ceiling', which were set to 60 and 75 Hz, and 500 and 600 Hz for male and female speakers respectively. Since the F0 tracking occasionally results in unreliable outlier values, rather than expressing the F0 range as the measured maximum-minimum, we computed the 80-percentile range (i.e. 90th – 10th percentile). Prior to the analysis, F0 tracks were turned into PraatPitchTier objects, visually inspected and corrected manually for errors such as octave jumps or creaky voice. One utterance out of the 944 was whispered and so it was excluded along with its post-test counterpart.

For the Trained Group, a subset of 7 phrases (italicised in Table 1) was also analysed for the range of the nuclear pitch accent. These phrases were expected to be realised with a falling melody by the learners. Due to an error during the pre-test administration, only 4 phrases were available for the analysis in 2 speakers in the Trained group. A total of 212 phrases were included in this analysis. To analyse the pitch range of the nuclear accent, the most prominent (accented) syllable was identified in each intonational phrase. If the prominent syllable had a clear local F0 peak, the peak was annotated as the F0 maximum. In the following unaccented syllable, the F0 minimum was chosen. If no clear F0 peak could be located and the F0 contour was flat, the stressed syllable in the word most likely to be in focus was marked as bearing the F0 maximum. Figure 2 illustrates the annotation with the phrase "*I think I'd better go now*," realised by the same learner on the pre-test (A) and the post-test (B). In (A), the prominent syllable "*go*" is marked by the local F0 peak while in (B), no clear F0 peak could be identified. The F0 maximum and F0 minimum were sometimes realised within the span of the same syllable when the pitch accent was assigned to the last monosyllabic word (phrases 8, 11, 13, and 14). In 3 recordings the prominent focus was impossible to determine and so they, and their counterparts from the pre- or post-test, were excluded.

Statistical analyses were conducted, and the plots were made in R (version 4.2.0, R Core Team, 2022) using the packages *lme4* (version 1.1.29, Bates et al. 2015), *ggeffects* (version 1.1.2, Lüdecke, 2018), *ggplot2* (version 3.3.6, Wickham, 2016) and *afex* (version 1.1.1, Singmann et al., 2022).

Table 1. The Stimulus Phrases and their Number of Syllables.

Phrase	Sylls.	Phrase	Sylls.
1. I wonder who that can be.	7	9. Excuse me, …	3
2. It can't be the milkman, …	6	10. … *but I am very hungry.*	7
3. … because he came this morning.	7	11. *Thank you for my nice tea.*	6
4. And it can't be the boy from the grocer, …	10	12. *I think I'd better go now.*	7
5. … because this isn't the day he comes.	9	13. *I don't know what to do.*	6
6. And it can't be daddy,	6	14. *The tiger has eaten it all.*	8
7. … because he's got his keys.	6	15. I know what we'll do.	5
8. *We'd better open the door and see.*	9	16. *I've got a very good idea.*	9

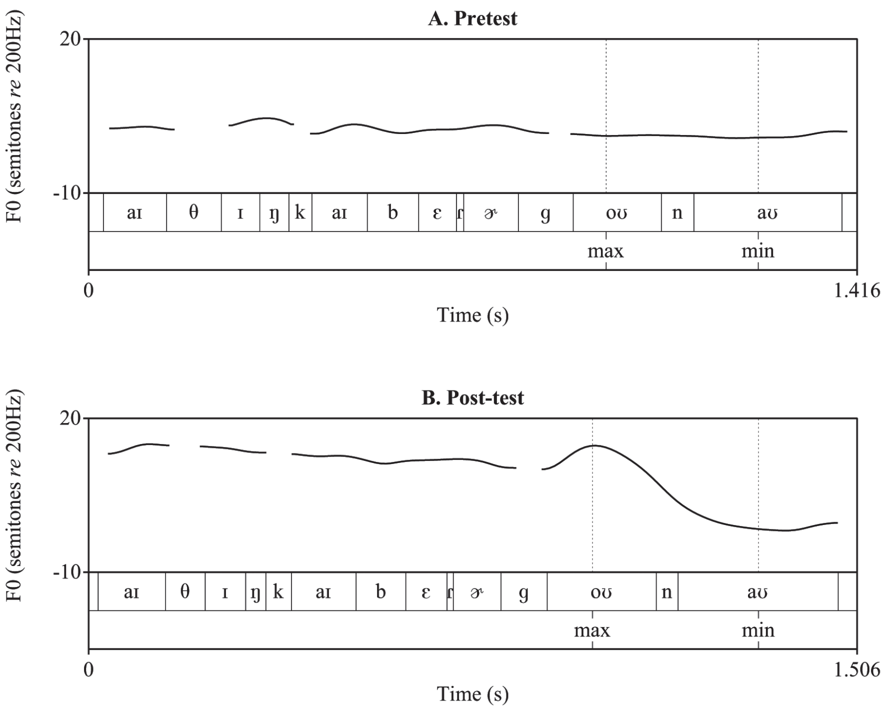

Figure 2. Example Annotation of the Pitch Range (F0 maximum and F0 minimum) of the Nuclear Accent

3. Results

3.1 Reading tempo

Figure 3 shows the distribution of the reading tempo values, expressed as syllables per second, across stimulus phrases and split by participant Group and testing Time. The figure suggests that while the speakers in the Comparison group hardly changed their reading tempo, there seems to have been a slowing down from Time 1 to Time 2 for many speakers in the Trained group.

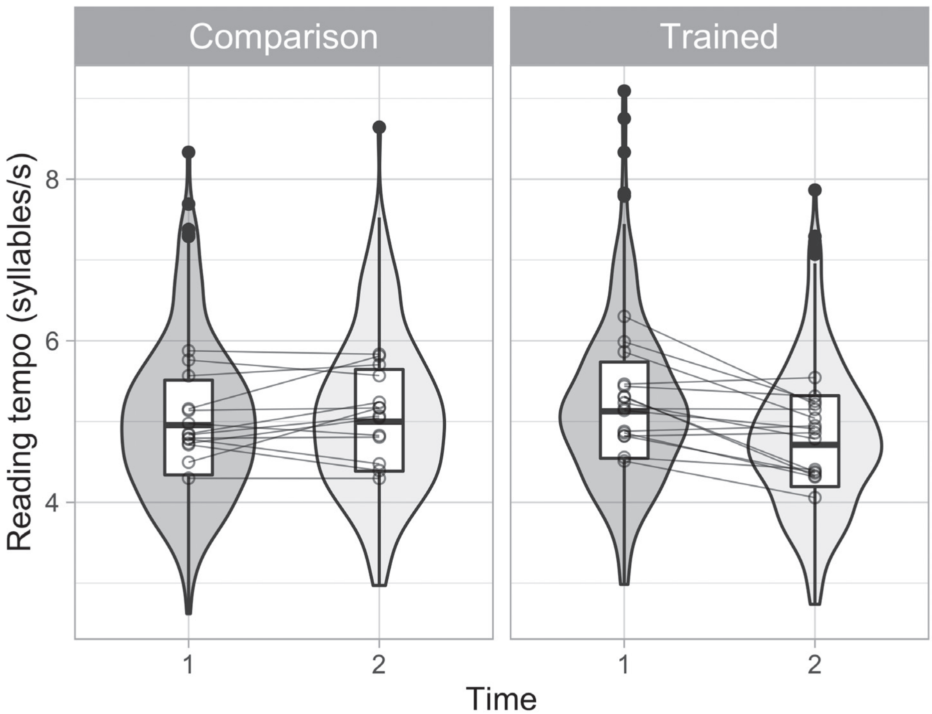

Figure 3. Reading Tempo at Time 1 and Time 2. Violin plots and boxplots show the reading tempo values (in syllables/s) across phrases split by Group and testing Time. Unfilled circles connected by lines show each speaker's means per Time.

To assess whether this difference was reliable, we fitted to the reading tempo data a linear mixed model with testing Time (Time 1 coded as 0, Time 2 as 1), Group (Comparison coded as 0, Trained as 1) and their interaction as fixed effects, and Participant and Phrase as random effects, each with varying intercepts and slopes for Time. Table 2 gives the estimated coefficients and Figure 4 plots the predicted reading tempo values. The model confirmed that whereas the tempo did not change reliably for the Comparison group (Time2 slope +0.09 syllables/s, $p > 0.3$), it did change for the Trained participants, with a predicted decrease of tempo from Time 1 to Time 2 by 0.09 – 0.49 = -0.4 syllables/s ($SE = 0.12$, $t = -4.03$, $p < 0.001$).

Table 2. Coefficients Estimated by a Linear Mixed Model Fitted to the Reading Tempo Data.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	4.9975	0.2156	30.7738	23.1809	**<0.0001**
Time2	0.0899	0.0921	29.8206	0.9766	0.3366
Group2	0.2195	0.1730	29.0724	1.2686	0.2147
Time2:Group2	–0.4915	0.1220	29.2039	–4.0289	**0.0004**

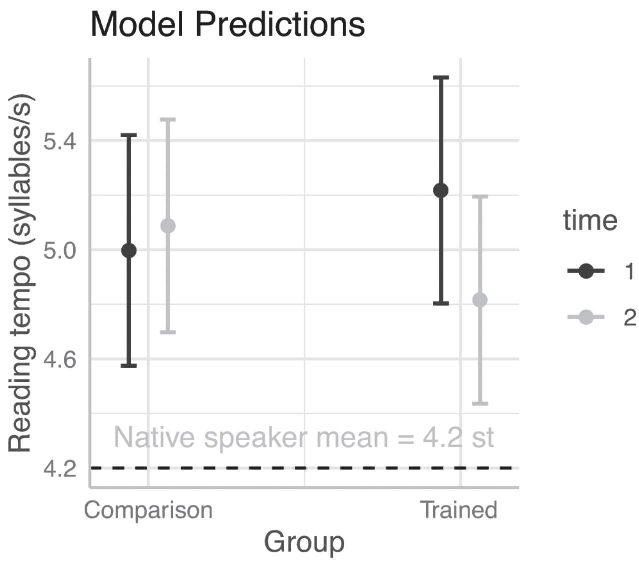

Figure 4. Predicted Reading Tempo Values with a Native-speaker Reference.

3.2 Pitch movement magnitude

As a measure of the magnitude of pitch movement, we used the F0 difference between the 90th and 10th percentile (see 2.3). The values measured, split by Time and Group, are plotted in Figure 5. The plot suggests that whereas there was no change in F0 range in the Comparison group between the testing times, for the Trained group there seems to have been an increase.

To determine whether this is a statistically reliable difference, we fitted to the data another linear mixed model, again with testing Time, Group (both again treatment-coded) and their interaction as fixed effects, and Participant and Phrase as random effects, each with varying intercepts and slopes for Time. Table 3 gives the estimated coefficients and Figure 6 plots the predicted 80-percentile F0 range values. The model confirmed that while the F0 range did not change significantly for the Comparison group (Time2 slope -0.13 semitones, $p > 0.8$), it did change for the Trained participants, with a predicted increase of F0 range from Time 1 to Time 2 by -0.133 + 2.736 2.6 semitones ($SE = 0.77$, $t = 3.53$, $p = 0.0013$).

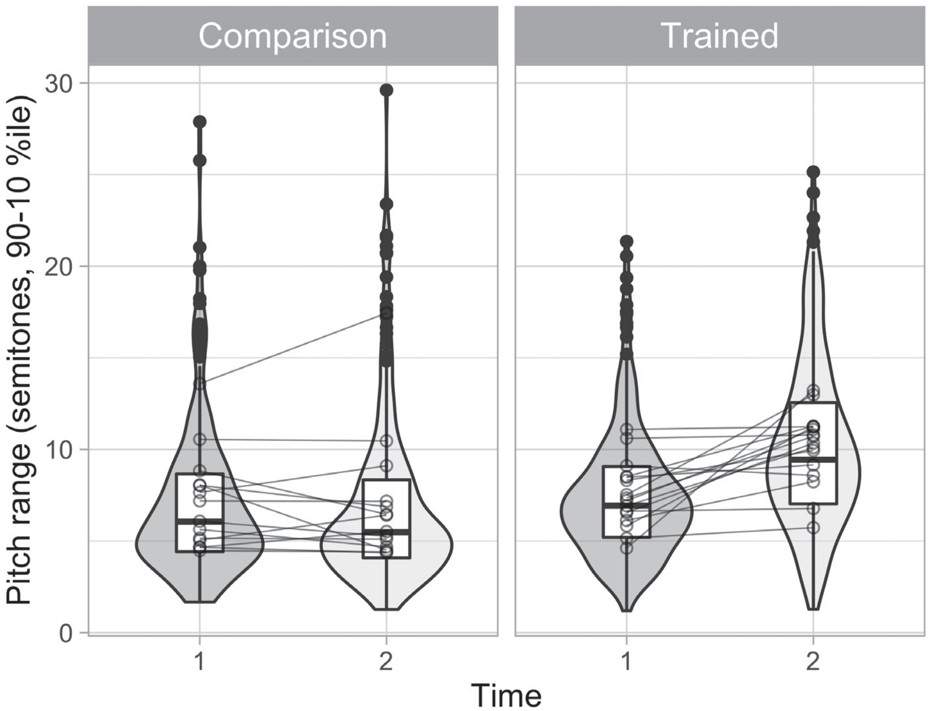

Figure 5. Pitch Range at Time 1 and Time 2. Violin plots and boxplots show the pitch range, measured as the difference of the 90th and 10th percentiles in semitones (re 200Hz), across phrases split by Group and testing Time. Unfilled circles connected by lines show each speaker's means per Time.

Table 3. Coefficients Estimated by a Linear Mixed Model Fitted to the 80-percentile F0 Range Data.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	7.1205	0.6431	38.4755	11.0715	**<0.0001**
Time2	–0.1332	0.5724	29.5157	–0.2327	0.8176
Group2	0.3881	0.7892	29.9104	0.4918	0.6265
Time2:Group2	2.7361	0.7733	30.0927	3.5381	**0.0013**

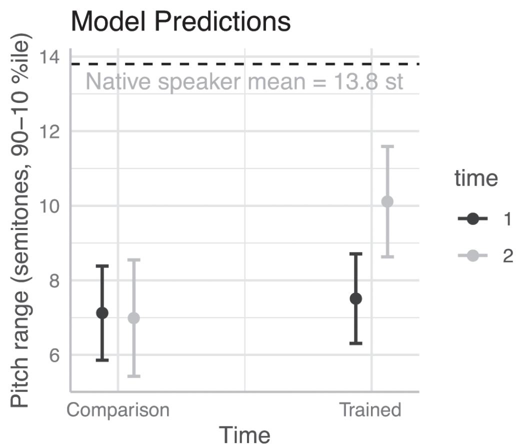

Figure 6. Predicted 80-percentile F0 range values with a Native-speaker Reference.

Next, for the Trained Group we measured the magnitude of F0 movement within nuclear accent contours (see 2.3 above for measurement details). Figure 7 plots the distribution of tonic accent range values for the two testing times, clearly suggesting an increase between times.

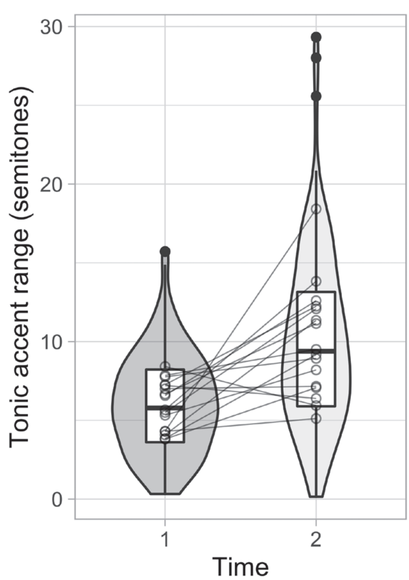

Figure 7. The Nuclear Pitch Accent Range. Violin plots and boxplots show the nuclear pitch accent range, measured as the difference in semitones (re 200Hz) between the F0 maximum in the accented syllable and the F0 minimum in the following unaccented syllable (see more details above in 2.3), across phrases split by testing Time. Unfilled circles connected by lines show each speaker's means per Time.

A linear mixed model fitted to the nuclear accent range data, with testing Time as the fixed effect and Participant and Phrase as random effects, each with varying intercepts and slopes for Time, did not converge. Thus, we refitted this model by restricted maximum likelihood (REML), along with a second model, again by REML, with a reduced random effect structure (see Matuschek et al. 2017 for a criterion for selecting random effects structure supported by the data), namely excluding the correlation of by-Phrase varying intercepts and slopes. A likelihood ratio test (LRT) found no significant difference between the two models ($p > 0.99$), indicating that dropping the correlation parameter did not significantly decrease likelihood. However, when refitting the second model by maximum likelihood, the model again failed to converge. We repeated the procedure, dropping the by-Phrase varying slopes in a third model, resulting in no significant reduction of goodness-of-fit as shown by a LRT ($p > 0.5$), but when the third model was refitted by maximum likelihood, a singularity issue occurred: a correlation between by-Participant varying intercepts and slopes equal to -1 was reported in the output. Thus in a fourth model, the correlation between by-Participant intercepts and slopes was excluded, leading to no decrease in likelihood ($p > 0.99$) but still the output reported a singular fit. A fifth model then removed the by-Participant slopes, this time leading to a significant decrease in likelihood ($p < 0.0001$). Therefore, we refit the fourth model by maximum likelihood, and we report the estimated coefficients in Table 4 (the note recapitulates the final random effect structure). This model on the data from the Trained group predicted the tonic accent range in Time 2 to be about 4 semitones higher than in Time 1 ($SE = 0.91$, $t = 4.30$, $p = 0.0003$).

Table 4. Coefficients Estimated by a Linear Mixed Model Fitted to the Tonic Accent F0 Range Data. Coefficients estimated by a linear mixed model fitted to the nuclear accent F0 range data with Time as the fixed effect and Participant (with varying intercepts and slopes, without slope/intercept correlation) and Phrase (with varying intercepts) as random effects.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.0553	0.4918	14.299	12.313	**<0.0001**
Time2	3.9526	0.9196	23.1834	4.2981	**0.0003**

Since the three models on reading tempo, F0 range, and nuclear pitch accent range, as reported in this section, were fitted to data collected from the same participants, it was necessary to adjust the alpha level from 0.05 to 0.05/3 0.016 (Bonferroni correction). All the *p* values reported as significant above are lower than that.

4. Conclusion

We examined the effectiveness of a semester-long general pronunciation course in improving prosodic skills of adult EFL learners. The course methodology is based on enhancing both perception and production skills, shifting zoom between phonetic detail and communicative effectiveness, and engaging learners in working with peer-to-peer as well as instructor-to-learner feedback. The study tested the utility of this methodology in the context of distance learning. The pre-to-post-test comparison focused on the learners' ability to read with prosody, specifically with the appropriate tempo and a wide pitch range. For advanced L2 learners who use English in professional capacities of teachers or interpreters, this is a useful skill.

Regarding the reading tempo, we expected that improvement in expressive reading would be reflected in a slower reading pace. This is what we saw in the post-test, though not uniformly for all trained learners. Closer inspection of the data suggests that duration gains are especially noticeable in the time given to words in focus; a future study should therefore include duration measurement of the pitch movement in the nuclear accented syllable. We also noted the number of disfluent utterances to drop from 24 to 5 in the Trained group, while remaining roughly the same in the Comparison group. It seems that the learners gained fluency as well as expressivity.

The learners also benefited from the training in terms of pitch movement. Clearly, the narrow pitch range regarded as typical for Czech-accented English can be expanded by guided practice even with adult learners who do not have the benefit of authentic input in an English-speaking environment. We measured improvements in the production sentential focus: the nuclear accent had a wider F0 movement in the Trained group's post-test. However, the current study does not consider other aspects of the F0 contour, e.g. pre-nuclear pitch accent, and the alignment of accents and segments.

We can conclude that the course helped the learners improve their L2 prosody in a reading aloud task. Like the example learner in Figure 1, most participants read more slowly and varied their pitch after the training. This was the case despite the training taking place online. While in-class learning has obvious advantages, as learners participate in actual rather than imagined interactions, receiving immediate rather than delayed feedback, the distant training did translate into pronunciation gains. One feature of the online course edition deserves to be explored further: the impact of the learners processing self- and peer-recordings. Also, a future study should compare pronunciation gains from the course taught in the face-to-face vs the distant learning environment to evaluate the importance of genuine oral interaction.

Acknowledgement

This research was supported the Czech Science Foundation grant number 21-09797S.

REFERENCES

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Cowie, R., Douglas-Cowie, E., & Wichmann, A. (2002). Prosodic characteristics of skilled reading: Fluency and expressiveness in 8-10-year-old readers. *Language and Speech*, *45*(1), 47–82.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment assessment*. Cambridge: Cambridge University Press.
De Mareüil, P. B., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica, 63*(4), 247–267.
Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.
Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory into practice, 30*(3), 165–175.
Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL quarterly*, *22*(3), 473–492.
Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication*, *36*(1-2), 31–43.
Huang, L. F., & Gráf, T. (2020). Speech Rate and Pausing in English: Comparing Learners at Different Levels of Proficiency with Native Speakers. *Taiwan Journal of TESOL*, *17*(1), 57–86.
Kang, O. (2008). *Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness* (Doctoral dissertation, University of Georgia).
Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566.
Kerr, J., & McEwan, G. (2006). *The tiger who came to tea*. HarperCollins Childrens' Books.
Kuhn, M. R., Schwanenflugel, P. J., Elizabeth B. Meisinger, E. B. 2010. Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly 45*(2): 232–253.
Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.) *Pragmatics and prosody in English language teaching*, Springer. 25–40.
Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (27–44). Multilingual Matters.
Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26), 772. doi:10.21105/joss.00772.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315.
Mennen, I., & de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition*, *36*(2), 183–194.
Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*(2), 159–182.
Munro, M. J., & Derwing, T. M. (2001). Modelling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language acquisition*, *23*(4), 451–468.
R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language learning*, *62*(2), 595–633.
Singmann H., Bolker B., Westfall J., Aust F., & Ben-Shachar M. (2022). *afex: Analysis of Factorial Experiments*. R package version 1.1.1, https://CRAN.R-project.org/package=afex.
Volín, J., Poesová, K., & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, *13*(1), 107–123.
Volín, J., & Skarnitzl, R. (2010). Suprasegmental Acoustic Cues of Foreignness in Czech English. *Proceedings of Odyssey: The Speaker and Language Recognition Workshop Odyssey 2010*. Brno: VUT, 271–278.
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Online data resources

YouTube. (2020, May). https://www.youtube.com/watch?v=UGqFxN0V-CA
YouTube. (2020, May). https://www.youtube.com/watch?v=6Cba9CUQUD0
YouTube. (2020, May). https://www.youtube.com/watch?v=IUAr_bnXtKc (no longer accessible)
YouTube. (2020, May). https://www.youtube.com/watch?v=Lok2QdctuCg (no longer accessible)

YouTube. (2021, June). https://www.youtube.com/watch?v=4k87lP1GNrE
YouTube. (2021, June). https://www.youtube.com/watch?v=ggjemuhlmuM
YouTube. (2021, June). https://www.youtube.com/watch?v=RjcxZemAxNI

Šárka Šimáčková
Dept. of English & American Studies
Palacký University Olomouc
Olomouc, Czech Republic
E-mail: sarka.simackova@upol.cz

Václav Jonáš Podlipský
Dept. of English & American Studies
Palacký University Olomouc
Olomouc, Czech Republic
E-mail: vaclav.j.podlipsky@upol.cz