CONDUCTING CLASSROOM RESEARCH: CZECH STUDENTS AS ENGLISH PROMINENCE AND MELODY ANNOTATORS

Alexey Tymbay

Abstract

The study posits that mid-sized groups of phonologically trained non-native speakers of English can collect prosodic data that are equivalent to English native-speakers' annotations. The hypothesis is supported by the results of a classroom experiment involving an experimental group of English-proficient Czech (L1) learners annotating prominence and boundaries in English monological texts before and after additional phonological training aided by Rapid Prosody Transcription (RPT). The annotation results received before the experimental group had the training demonstrate deficiencies of their prosodic annotation occurring under the probable influence of the learners' mother tongue (Czech). The analysis of disagreements between the experimental group's and the control group's (native speakers) annotations demonstrates that non-native listeners rely on slightly different cues when identifying the prosodic structure of an English utterance. Thus, it is concluded that Czech (L1) speakers of English require mandatory annotation practice focused on the differences between their mother tongue and English to perform annotation tasks successfully. The experimental group's RPT annotations, conducted after a learning intervention, produced much better results and were recognized as statistically equivalent to native speakers' RPT annotations. The high alignment of the readings obtained by the experimental and control groups on key prosodic parameters demonstrates that crowdsourcing prosodic information from phonologically trained non-native speakers with the help of the RPT method can be employed as an alternative means of validating intonation research when attracting native speakers to research participation is problematic.

Keywords

prominence, boundaries, prosodic interference, Rapid Prosody Transcription, auditory analysis

1 Introduction

The established procedure of experimental phonetic research conducted by non-native speakers of English typically requires its results to be externally validated by native speakers of English (Demolin 2012, Spreafico & Vietti 2022). Although English today is more commonly used as a lingua franca, their judgments and perceptions are still considered more reliable indicators of how speech is perceived within the British or American language community (Baese-Berk et al. 2020). Therefore, to ensure that the findings of phonetic research are valid

Discourse and Interaction 17/2/2024, pp. 115-146
ISSN 1802-9930
https://doi.org/10.5817/DI2024-2-115

and generalizable, it is still customary to include native speakers in the validation process, with their input helping establish that the study accurately represents the phonetic patterns of the language norm often understood as a pronunciation standard in a particular geographical area.

Indeed, native speakers of a language are believed to have a deeper understanding of the nuances of pronunciation and phonetics, which may not be apparent to non-native speakers, as they "perceive speech not only in terms of the acoustic signal but also with their own experience and biases driving their perception" (Baese-Berk et al. 2020: 1). Native listeners typically grasp their language's prosody intuitively to identify when pronunciation deviates from the expected standards (Kirkova-Naskova 2010). They are also well-equipped to provide insights into phonetic variation's sociolinguistic and cultural aspects within their mother tongue, as "sociophonetic variability is strongly governed by the individual's exposure to the statistical properties of ambient sound patterning and […] social context" (Foulkes & Docherty 2006: 433).

On the other hand, Baese-Berk et al. (2020: 13) insist that "our understanding of language processing can no longer be limited to the use of a native, monolingual English speaker as 'the ideal speaker-listener'". In the researchers' opinion, for a phonetic study to be representative of natural language processing, it should investigate a broader range of speakers, listeners, social factors, and communicative situations. However, non-native listeners' ability to adequately assess the key prosodic features of a foreign language still raises many concerns, as non-native listeners may unintentionally misrepresent or misinterpret certain phonetic phenomena due to their limited exposure to the nuances of native speech and interlanguage barriers. Additionally, Jenkins (1995: 34) maintains that the unconscious transfer of L1 language patterns may create perceptual "substitutions" to the critical elements of target language phonology, "which persist in interlanguage the longest" and may significantly distort the correct perception of both segmental and suprasegmental cues.

Smith and Edmunds (2013: 235) report that to evaluate the impression that speech makes on listeners with various native languages, Rapid Prosody Transcription (RPT) may be a most valuable method, as "it reflects listeners' adaptation to different speakers". Indeed, RPT effectively records listeners' immediate perception of prosodic elements calculated as the p-score (prominence score) and b-score (boundary score) for each word in the transcript, thus expressing the inter-transcriber agreement. According to the researchers, the involvement of non-expert and non-native transcribers in RPT annotation tasks can further enrich data on foreign language prosody perception, as it allows viewing cross-transcriber differences as a source of information about the areas where challenges of prominence and boundary perception may arise.

Taking into account that cross-influence of every pair of languages (learner's L1 and L2) can have an unpredictable impact on the perception of the suprasegmental features of the target language due to the inherent L1 transfer, more studies are needed to explore the effects of this interference and assess the validity of learners' perceptual findings. This paper underscores the need to study Czech learners' perception of English prosodic features (prominence and boundaries) and compare the results of their auditory analysis with native-speaker perceptions. Key prosodic differences between Czech and English expressions of prominence should also be considered to assess the possible discrepancies in the results thoroughly.

This paper hypothesizes that after specific phonological training, primarily based on avoiding the prosodic interference of the Czech language, English-language-proficient Czech (L1) speakers employing the RPT method may provide accurate prosodic data about an English monologue comparable to native speaker results.

The study was conducted to find answers to the following research questions:

1. How do the results of an auditory RPT analysis performed by Czech learners of English compare with the auditory data from native English speakers?
2. Does Czech speakers' perception of English prominence and boundaries change after additional phonological training?
3. What aspects of Czech-to-English prosodic interference should be highlighted in class to minimize the possible impact of L1 intonation transfer?

2 Literature review

2.1 The role of students' native languages (L1) in target language (L2) prosody perception

For many decades, phonological research has underscored the role of native language in the perception and production of L2 phonetic features. For instance, Trubetzkoy (1939) was one of the first researchers to suggest that the phonological system of the mother tongue acts like a sieve that passes only the phonetic information considered relevant in the student's native language. However, the degree of the impact that L1 prosodic features have on L2 perception still needs researchers' attention as functions, meanings, and realizations of suprasegmental elements are different in every language. Moreover, new approaches to L2 acquisition have indicated that the relationship between L1 and L2 phonology can be not as straightforward as previously thought. So, it is necessary to study crosslinguistic influence in each particular pair of languages and consider various interpretations of the results.

Contrastive analysis of phonological systems maintained that the ease or difficulty of a specific L2 pronunciation feature could be predicted through the "exhaustive analysis" of its phonetic details in the learner's mother tongue (Brière 1966: 795). Guided by the contrastive analysis traditions, Mennen (2015) suggests that by comparing the learner's L1 and L2, it is possible to predict where deviations in L2 perceptions resulting from inevitable L1 interference may occur. In the context of L2 Intonation Learning theory (LILt), Mennen (2015: 173) proposes a model for comparing suprasegmentals, which encompasses four dimensions: 'systemic', 'realizational', 'semantic', and 'frequency' that help to identify the right approach to L2 intonation instruction. According to the researcher, this model of intonation comparison offers "all necessary tools for a thorough characterization of intonation deviation", which is to be addressed through targeted perception and production practice (Mennen 2015: 178).

Alternatively, Eckman's *Markedness Differential Hypothesis* (1977) stated that the areas of difficulty that an L2 learner would have could be predicted based on a systematic comparison of 'marked' and 'unmarked' phonological features in their L1 and L2. However, in criticism of this hypothesis, Hume (2004: 193) posits that "given standard assumptions, markedness only predicts the patterns that are supposed to be universal; it does not provide predictions regarding language-specific markedness patterns". The researcher insists that markedness should be used only as a descriptive instrument rather than an ultimate explanation of a feature's relative difficulty for perception since "the expectations of language users and predictability of the elements differ considerably" (Hume 2004: 193).

Follow-up studies that focused on non-native listeners' perception of prosodic features have indeed proved that "non-native speakers do not fully control the cues used by native English speakers to indicate prosodic patterns" (Smith & Edmunds 2013: 235). Van Maastricht et al.'s (2017: 367) attempts to explain this difficulty reduced the factors inhibiting learners' perception of L2 intonation to "transplanting donor intonation", understood as a blind transfer of L1 patterns to L2 perception, and "insufficient proficiency levels" of the listeners. The researchers also recognize that "intonation is highly context-dependent", so it might be difficult to make relevant phonological distinctions "based only on the acoustic cue without any context at all" (ibid.) as "sensitivity to low-level acoustic input is modulated by higher top-down knowledge of the language" (Grice & Kügler 2021: 255).

Schmidt et al. (2016: 245) also demonstrate that L1 experience "shapes the perception of suprasegmental information". However, they note that pitch processing at a word level occurs independently of pitch processing at an utterance level. This means that general auditory sensitivity, evidently tied to

the L1 experience, and perceptual sensitivity to suprasegmental information are not the same. Thus, it is concluded that an increase in the proficiency levels, "requiring extra coaching and specific training to improve intonation, might be a useful addition to an L2 curriculum" (van Maastricht et al. 2017: 367) if a tangible shift in foreign language intonation perception is expected from a learner.

To assess how relevant these ideas are in foreign language classrooms, Lee et al. (2020) investigated whether phonetic training focused on perception or production is more effective in improving the English learner's pronunciation skills, with their findings indicating that phonological training emphasizing perception is more beneficial than production-focused instruction probably because, as noted by Trofimovich and Baker (2006: 23), if learners are not able to perceive "(subtle) crosslinguistic differences" in the input they receive, they fail to "align the relevant prosodic cues with their phonetic realizations". Although Lee et al. (2020) recognize that a person's perceptual system gradually becomes exclusively attuned to L1-specific features and thus progressively worse at discerning elements that are not contrastive in their L1, the researchers, nevertheless, conclude that "guiding students to improve their perception rather than production ability may maximize the process and product of acquisition under classroom conditions" (ibid.: 3).

Considering this evidence of the benefits of extensive L2 exposure on target language phonology as a primary factor, it is also important to note the role of additional phonological instruction in students' discrimination ability. The meta-analysis of earlier studies allowed McAndrews (2019: 156) to conclude that "instruction for suprasegmental features can benefit L2 learners who have previously had substantial exposure to the target language". McAndrews' conclusions show that even brief instructional sessions on L2 prosody, including those lasting less than three hours in total, produce a significant and positive impact on learners' ability to recognize and categorize suprasegmental features of the target language, irrespective of their proficiency levels.

Although McAndrews (2019: 158) mainly focuses on the additional suprasegmental instruction as the best way "to build comprehension", the author also extends the positive effects of it to "expert listeners" who can consequently learn from this instruction to note the role of suprasegmental features in a diverse range of essential functions "from recognizing words to parsing sentence structure and taking conversational turns". This generalization is significant for our case study as it helps to substantiate our assumption on the potential of advanced L2 learners of English who had undergone targeted phonological training to be used as subjects in prosodic research.

To conclude, various studies of second language acquisition (the term 'second language', or L2, is treated in this paper as a synonym of a 'foreign language', or FL) uniformly agree on the fundamental role that native languages play in the learners' perception of L2's phonological features, with 'positive transfer' (easy recognition of the target language's marked features happening in the absence of fundamental systemic, realizational, or semantic differences between L1 and L2 phonological systems) and 'negative transfer' (overlooking the marked features of the target language because of the native language influence, e.g. frequency or realizational differences) being the two possible varieties. The studies also confirm that the adverse effects of native language interference can be potentially eliminated by increasing exposure to the target language, as "the accuracy with which L2 phonetic elements are perceived and produced depends on how much native-speaker input is received" (Flege et al. 2003: 467) or additional phonological instruction, which is believed to critically improve the L2 comprehension even after short periods of targeted training.

2.2 Rapid Prosody Transcription (RPT) for prominence and boundary annotation

Since, according to Trofimovich and Baker (2006), the effect of additional suprasegmental training is dependent on the particular suprasegmental feature under discussion, a decision was made to narrow the scope of the present study to the perception of English prominence and boundaries by Czech learners of English. Apart from the widely recognized importance of prominence on language intelligibility, prominence is crucial to us in the way that it combines two levels of auditors' perception: 'down-up' – the acoustic cues of the speech signal – and 'top-down' – the linguistic context of the utterance (Bishop et al. 2020). In their recommendations for prosody studies, Cole et al. (2017: 310) make a list of the acoustic and contextual factors potentially affecting prominence perception (Figure 1).

As Cole and Shattuck-Hufnagel (2016: 5) report, in the perception of prominence "the association between a prosodic element and its function involves not a single cue but a constellation of cues, with the cue mixture and cue values subject to variation". In natural speech settings, listeners integrate these features to identify prominence based on "the entire cue package" and "the relationships between neighboring cues"; however, due to the complexity of acoustic marking, combined with individual speaker differences and various contextual cues, ambiguities in prominence perception, even among phonologically trained native speakers, are not uncommon (Cole et al. 2017: 300).

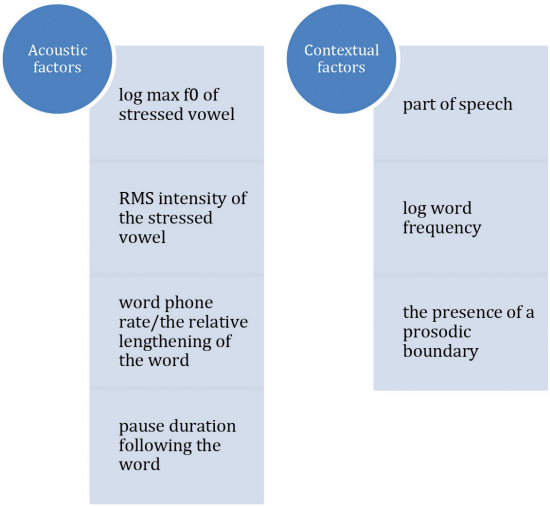

Figure 1: Acoustic and contextual cues for prominence and boundary marking

Since neither the traditional British annotation system nor the ToBI (Tones and Break Indices) convention provides simple-to-use yet reliable labels to mark prominence, Cole and Shattuck-Hufnagel (2016: 1) suggest using RPT – "a simple set of unary labels to mark prominence and boundaries based on immediate auditory impression" – as a means of attracting larger audiences to prosodic data collection. In the RPT method, inter-transcriber agreement in small-to-medium experimental groups of auditors is used to calculate the prosody "score" for each word or boundary, and inter-transcriber differences are viewed as a source of information about where ambiguities are likely to arise (Cole & Shattuck-Hufnagel 2016: 7-8).

The procedure of RPT annotation is the following: transcribers listen to recorded samples through headphones and are asked to mark prominence and boundaries in a specially prepared transcript (all punctuation marks and capital letters are removed) of the speech sample. The results are processed manually or digitally aided by an open-source software tool LMEDS (Language Markup and Experimental Design) (Mahrt 2016) to calculate the word's respective 'p-score' (prominence score) and 'b-score' (boundary score). The scores are calculated as a ratio of the number of participants who marked the respective feature to the total number of annotators (cf. an abstract of an RPT-annotated transcript in Example 1, with the examples below coming from the Czech learners' annotations

obtained during the study). The figure close to 1 indicates the presence of a perceived boundary/prominence in the recorded sample, whereas the figure close to 0 indicates the annotators' uncertainty (Cole & Shattuck-Hufnagel 2016).

(1) *on this **wonderful**(.36) **day**(1) |(.83) when we are **gathered**(.33) **together**(1) |(1) to celebrate your academic **success**(.92) |(.62) I have **decided**(.83) |(1) to **talk**(.83) to you |(.56) about the **benefits**(.42) of **failure**(.92) |(1) and as you stand on the **threshold**(.33) of what is sometimes called **real**(.66) **life**(.92) |(1) I want to **extol**(.83) the **crucial**(.42) importance of **imagination**(1) |(1)* – The annotators marked prominences as **bold** and put breaks (|) where they perceived boundaries. P-scores and b-scores (in brackets) for each prominence and boundary were then calculated by dividing the number of annotators who marked the respective feature by the total number of annotators.

The included in RPT marking of boundaries facilitates further categorization of identified prominences into two main types: nuclear-accented – the prominences preceding a phrasal boundary – and pre-nuclear accented words – occurring without a subsequent boundary. Typically, pre-nuclear accents initiate a tone group, defined as a group of words between the two boundaries, while nuclear accents conclude it (Example 2). Nuclear tones are essential to every tone group, whereas pre-nuclear accents are discretionary and play a rhythmical role.

(2) *| a **language** | becomes a **global** language | because of the **power** of the people who **speak** it |* – The utterance consists of three tone groups/units separated by phrasal boundaries (marked "|"). Each tone group/unit, understood in this paper as a group of words demarcated by salient interruptions to a flow of speech (e.g. silences, filled pauses, or perceived breaks without a silent interval) comprises an obligatory nuclear accent ("language", "global", and "speak"). The final tone group also contains a pre-nuclear accent ("power") identified as such because it is not followed by a boundary but another accent.

According to Bishop et al. (2020: 16), apart from its proclaimed simplicity, another advantage of using RPT for prominence and boundary annotation lies in the fact that it allows for an integrated perceptual assessment of prosodic elements irrespective of the exact way listeners identify them (either "down-up", that is guided solely by acoustic cues, or "top-down", lead, for instance, by the word's informativeness). As a result, RPT provides phoneticians with simple-to-use prosodic labels that demonstrate "how information from the continuous speech signal gets translated into information about the linguistic structure, information structure or discourse reference" (Cole & Shattuck-Hufnagel 2016: 6).

In contrast to other annotation methods, where experts aim to achieve a certain level of agreement on their prosodic readings, the perceptual weighting of elements in RPT aims at pointing out the areas of perceptual ambiguity, which, in our experimental study of Czech learners' ability to annotate English prominence, may assist in identifying the differences in the perceptual cues used by native and non-native listeners. Furthermore, apart from being "an efficient means of transcribing prosody in spontaneous speech produced in interactive communication contexts" (Cole & Shattuck-Hufnagel 2016: 9), RPT annotation tasks seem to be an effective teaching tool employed in a language classroom for enhancing the understanding of prosody's role in foreign language comprehension.

2.3 The problematic areas of Czech-to-English intonation transfer

As described above, the perception of L2 prosody is tightly connected with annotators' native languages, and the transfer of annotators' L1 perceptual habits to English may seriously conflate their perception of English prominence and boundaries. Researchers believe that "lack of differentiation might result in marking either far more or far fewer prominences" (Smith & Edmunds 2013: 237); however, focused phonological instruction can radically improve the comprehension and discrimination of the target language's prosodic features (McAndrews 2019). A preparatory training course, therefore, has to consider the systemic, semantic, realizational, and frequency differences in prominence and boundary characteristics. The following review compares and contrasts the critical characteristics of functions and realizations of English and Czech prominence with the key takeaways summarized in Table 1 below.

We acknowledge the importance of the distinction between word stress and sentence stress in prominence identification. However, as Cho and Keating (2009: 468) report, "stress and accent represent degrees along a single scale of prominence, and thus manifested in the same set of physical properties". Therefore, we believe the comparison of prominence patterns in L1 and L2 would be incomplete unless the word stress patterns of the mother tongue are taken into account, as both word stress and sentence stress can be presented as co-related points along a continuum of prominence, sharing common acoustic attributes. Acoustic dimensions such as F0, duration, intensity/spectral tilt, and vowel quality are utilized in both lexical stress and phrasal accent, linking their perception across various languages. Even though some phonetic research excludes lexical tone and word stress from prosody, they still exhibit significant parallels with typical prosodic features, particularly in language teaching contexts.

	The English language	The Czech language
The characteristics and functions of word stress:	There is no fixed word stress position. Word stress has a phonological (contrastive) function but no delimitating function.	Word stress is fixed on the first syllable; it plays no phonological role but has a delimitating function.
The relation between word order and prominence:	A fixed word-order language; acoustic marking of pre-nuclear and nuclear prominence is observed.	A relatively free word-order language; a shift of contextually important words towards the end of the utterance is observed.
The characteristics of prominence:	Prominence is signal-driven, i.e. it primarily relies on the word's prosodic marking.	Prominence is expectation-driven, has limited acoustic marking, and is mainly context-based.
The acoustic marking of a stressed syllable/ word:	A stressed syllable is marked by a peak F0, extended pitch range, vowel lengthening, and (often) higher intensity.	A stressed syllable has limited acoustic marking: lower F0, no vowel length change, and no intensity change is observed.
The acoustic marking of unstressed syllables/words:	The unstressed vowels significantly reduce their length and quality, i.e. 'centralized'.	Unstressed vowels are not reduced. Final vowels tend to be longer.

Table 1: The notable differences between English and Czech stress/prominence based on the literature review in Section 2.3

Volín and Zimmermann's (2011) studies of the acoustic nature of Czech stress, which is fixed on the first syllable (a systemic difference with English), discovered significant discrepancies in the duration and spectral tilt of stressed syllables. This absence of consistency downplays the role of acoustic cues in Czech stress marking while emphasizing individual spectral characteristics of every separate syllable type non-transferrable to a foreign language. In contrast, consistent acoustic changes, such as the stressed vowel positional lengthening, a distinct pitch change and a potential rise in intensity, are reported essential for marking stress in English (Adams & Munro 1978), with the positional vowel lengthening being the most evident cue for detecting the stressed syllable (a realizational difference).

The same findings can be partly extrapolated to the acoustic marking of prominence. Comparative studies of Czech and English prosody (Volín et al. 2015, Skarnitzl & Hledíková 2022) conclude that prosodic signaling is relatively weak in Czech, "with the stressed syllable bearing no typical signs of acoustic prominence, prosodic phrases being much longer than in English and pitch range being considerably narrower" (Čtvrtečková et al. 2023: 33). In a contrastive English vs. Czech prominence study, Weingartová et al. (2014: 239) conclude

that British English native speakers "exhibit distinct patterns [of prominence marking] compared to Czech respondents across all the parameters". The acoustic attributes of the British English speakers in the experiment align with the conventional prominence patterns documented in the literature: "stressed vowels among English native speakers displayed longer duration, higher SPL, a greater F0, and a flatter spectral slope compared to unstressed vowels" (Weingartová et al. 2014: 239). In contrast, Czech speakers do not consistently and comprehensively establish prominence contrasts to the same degree as their native-speaker counterparts (a realizational difference).

The absence of vowel reduction in Czech (a systemic difference) also significantly diminishes the presence of distinct phonological contrasts in Czech-accented English. Their role is misrepresented both functionally and acoustically: "Czech has no systematic vowel reduction, so vowels tend to be realized as full in their quality, even in unstressed syllables in Czech-accented English" (Čtvrtečková et al. 2023: 33). The natural phonological transfer of this phenomenon to English production and perception may thus significantly aggravate the identification of prosodic prominence in English by Czech native speakers.

Additionally, Chamonikolasová's (2018) study of Czech and English dialogues shows that the length of a standard tone group in Czech is much longer than in English and, consequently, the ratio of nuclear accents in the two languages is significantly different (a frequency difference). As the analysis of the final pitch movements suggests, both English and Czech declarative sentences predominantly exhibit a falling pitch direction (ranging from 69.7% to 95.0% of all terminal tones); however, Czech non-terminal units more frequently have a rising pitch (45.4% vs. 19.9% in English) than a falling pitch (34.0% vs. 53.3% in English) (ibid.: 66-70).

It means that although the universal prosodic variables in the two languages function similarly (e.g. a falling tone equally expresses the idea of completion in both languages), the differences in their acoustic realization, such as the tone group length or the rise/fall ratio, result in severe discrepancies in nearly all applications of intonation that foreign language students should be aware of. For example, the prevalence of rising tones in non-terminal units may signal the differences in discourse (turn-taking) function. A narrower pitch range, transferred from students' L1 (Czech), may cause attitudinal differences in tone perception. Mildly expressed stressed syllables, the absence of vowel reduction and the lack of marked nuclear tones weaken the delimitation and accentuation of Czech-accented speech. The cumulative effect of the reliance on wrong acoustic cues and the negative transfer of L1 prosodic habits to English may result in the

complete change of the stress-timed English rhythm to a syllable-timed effect that is more customary to Czech (L1) speakers (Tymbay 2022).

As mentioned, achieving near-native listening skills entails addressing positive and negative language transfer (Lee et al. 2020). To effectively mitigate Czech-to-English prosodic interference, attention should be directed towards specific areas that pose significant challenges in English prosody production and perception, particularly the systemic and realizational differences between the two languages. This involves highlighting the distinct prosodic features of English that contrast with Czech, such as the heightened acoustic prominence of stressed syllables in English, which includes variations like positional vowel lengthening or reduction, a broader pitch range, and a wider variety of tones. Conversely, aspects marked in Czech but unmarked in English, such as fixed stress on the first syllable and longer prosodic phrases, should be neutralized to minimize the transfer of the patterns to English prominence perception and production.

In conclusion, taking into account the differences between the prosodic expression of prominence and boundaries in Czech and English and the established transfer of learners' perceptual habits from their mother tongue (Czech), it is crucial to assess whether this anticipated negative impact on English perception can be minimized by targeted phonological instruction. Since RPT is generally considered a most helpful tool for probing non-expert and non-native perception due to its proclaimed simplicity and universality, this method of prosodic analysis seems most beneficial for our research purposes as it allows for comparing the perception of audio recordings at different levels of learners' phonological training, with disagreement analysis highlighting the challenges to non-native listeners' perception. Comparing the acquired prosodic data with native speakers' RPT impressions assists us in drawing conclusions about the quality of non-native prosodic data.

3 Method

The study was organized as a perceptual analysis of selected English monologues utilizing RPT, in which a relatively homogeneous group of Czech native speakers with a high English proficiency level analyzed English prominence and boundaries before and after a training course in phonology (Figure 2).

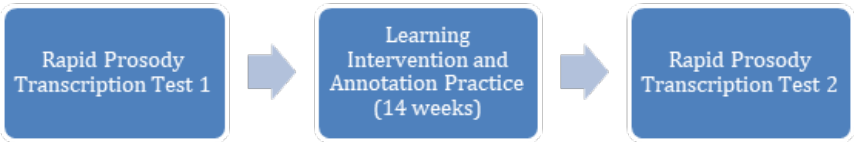

Figure 2: The experimental stages

The Czech annotators' RPT tests were eventually compared to RPT annotations performed by British English native speakers, who comprised the control group. Their boundary and prominence markings were conditionally accepted in this study as the ground truth.

3.1 Procedure

The experiment was conducted in two steps: Stage 1 (Week 1, before the start of the training course) involved an RPT (1) annotation task. This stage aimed to gauge Czech speakers' perception of English prominence and boundaries before they took a targeted phonological course. To this end, the participants were briefly instructed about the function of prominence (marking contextually important words), its acoustic nature (stressed words that stand out from the utterance, typically marked by increased loudness, length and a marked pitch change), and the essence of boundaries (perceptual breaks to the speech flow, meant to separate tone groups/ideas) in connected speech. They were provided with paper transcripts where they could mark breaks (|) and underline/circle the prominent words.

Following Bishop et al.'s (2020) recommendations for conducting prominence analysis, the annotators had only three passes (listenings) of the recordings (Stage 1) as a higher number of passes is reported to proportionately increase the number of false identifications. The participants were asked to familiarize themselves with the text (Listening 1), to mark the perceived breaks (Listening 2), and to mark all prominent words (Listening 3). The recordings were played to the annotators in a language lab equipped with individual headphones for every participant. The annotations were performed in real-time (the recordings were not paused for marking).

The same procedure was repeated in Stage 2 (Week 14, at the end of the training course), with the only difference being that the experimental group members were now given only two passes of the Stage 1 recordings as they were already familiar with the texts from the previous experimental stage. Stage 2 used the same recordings in the same order as in Stage 1 to avoid excessive variability of the research samples and ensure better comparability of the RPT results.

The control group (the native speakers) did the annotation separately on a different date and had three passes of the recordings. In contrast to the experimental group members, native listeners were allowed to pause the recordings and discuss their findings to create the ground truth data set required to validate non-native speakers' annotations. This collective experience was considered essential to mitigate the impact of individual perceptual differences on their prominence perception and get a consensus annotation against which the participants' annotations were compared.

3.2 Learning intervention

All participants of the current study underwent a specially developed one-semester course to minimize Czech prosodic transfer and enhance their English suprasegmentals' perception. SPELB (Experimental Phonetics) is an elective subject taught at the Technical University of Liberec (Czech Republic) focusing on English vs. Czech prosodic differences and methods of speech analysis. The fourteen-week course, including fourteen 90-minute-long seminars, was considered an indispensable element of the present experiment meant to identify the effect of additional intonation practice on target language perception. The key course components are presented in the Appendix.

The phonology course was developed following the list of major issues identified in the accented speech of Czech learners of English in earlier research (Skarnitzl & Rumlová 2019: 121-124, Tymbay 2022: 7-8). The course was also aimed at the mitigation of negative Czech-to-English intonation transfer, including:

- Eradicating inconsistent stress/prominence placement of Czech speakers of English and promoting the reduction of unstressed syllables/words in their English speech;
- Building students' melodic awareness and widening the range and repertoire of the tones used;
- Increasing learners' intonation awareness by working with different types of intonation annotations (RPT, traditional (British), ToBI) and Praat software (https://www.fon.hum.uva.nl/praat/) for the acoustic analysis of speech signals;
- Enhancing the rhythmic patterns of their speech, including focusing on prominent words' accentuation and contrasts between stressed/unstressed syllables.

3.3 Research participants

The experimental group: The Czech annotators in this study comprised twelve second-year BA students at the Technical University of Liberec studying English as their major at the Faculty of Science, Humanities and Education. The participants, consisting of nine female and three male students, had a proven level of proficiency in English (C1) and did not report any hearing problems. Their level of language proficiency was preliminarily checked by a computer-based level test (https://www.testgorilla.com/test-library/language-tests/english-proficient-c1-test/) and supported by the entry requirements for the SPELB course (the completion of a C1 focused course was a mandatory precondition).

Regarding annotation experience, the experimental group participants were initially considered amateur annotators (Stage 1). This categorization stemmed from their limited exposure to suprasegmental transcription methods in the introductory phonetic courses. However, intensive suprasegmental training, including rigorous auditory recognition practice, the ability to differentiate tones contrastively, imitation, and annotation practice conducted over the 14 weeks of the SPELB course, elevated them to the status of phonologically trained annotators (Stage 2).

The control group: The group of English native speakers was comprised of two experienced British lecturers (one male and one female) holding qualifications such as a PGCE (Postgraduate Certificate in Education), DipRSA TEFL (Teaching English as a Foreign Language), and LTCL DipTESOL (Diploma in Teaching English to Speakers of Other Languages). These lecturers were familiar with the RPT annotation system, obviating the need for additional phonological training.

We admit that the age difference between the experimental group participants (on average, 22-24 years old) and the control group (50-54 years old) could have potentially distorted the achieved results, as the listener's age is reported to significantly impact the perception of many acoustic parameters (Tremblay et al. 2003). Still, in the presence of a more notable distinction, namely the participants' native languages, for this study, the age difference between the control and experimental groups might be ignored and left for further research.

3.4 Materials

Four recorded monological abstracts from public speeches were selected for annotation analysis. Excerpts lasting c. two minutes each were extracted from four YouTube video clips, converted into WAV format to minimize extraneous noises and played to the students as sound files. All four speeches were categorized as Standard Southern British English (SSBE) and comprised two male and two

female speakers. The recorded samples can be classified as semi-spontaneous monologues delivered with a communicative intention to inform the audience. Detailed information about the recordings is presented in Table 2.

	Recording 1	**Recording 2**	**Recording 3**	**Recording 4**
Duration	1:56	1:42	2:40	3:03
Speaker	Joanne Rowling	David Cameron	Theresa May	Jeremy Hunt
Gender	female	male	female	male
Word count	295	291	406	388

Table 2: The recordings' characteristics

The transcripts of the recorded monologues were specially prepared for the experiment. To avoid visual guidance of the annotators, all capital letters at the start of grammatical sentences and punctuation marks were removed from the texts (Example 3). The prepared transcripts of the recordings are available from the publicly accessible repository at: https://github.com/Reading-between-the-lines/RPT-annotations.

(3) *on this wonderful day when we are gathered together to celebrate your academic success I have decided to talk to you about the benefits of failure and as you stand on the threshold of what is sometimes called real life I want to extol the crucial importance of imagination these may seem quixotic or paradoxical choices but please bear with me*

The original recordings can be found on YouTube:
Recording 1: https://www.youtube.com/watch?v=wHGqp8lz36c (3:10 – 5:03)
Recording 2: https://www.youtube.com/watch?v=WZI1EjxxXKw (0:19 – 1:58)
Recording 3: https://www.youtube.com/watch?v=FDyZ8trge2E (1:30 – 3:37)
Recording 4: https://www.youtube.com/watch?v=1N2yk0ubNaI (4:00 – 6:08)

3.5 Analyses

Quantitative analysis: Following Cole et al.'s (2017) guidelines for RPT results' verification, for medium-sized groups consisting of ten to twelve participants, the p/b-score values are not considered "sufficient to guarantee replicability of annotations at the group level" (Cole et al. 2017: 306) if the number of annotators who marked the feature is less than five. Therefore, a decision was made to establish a 50 per cent threshold for the p/b-score to consider the presence or absence of a respective feature in the utterance. For our experimental group consisting of twelve participants, a 50 per cent threshold corresponds to

a p/b-score =.6, calculated as the number of participants marking the respective feature (min. six) divided by the total number of participants (twelve).

The prominent words which gained a higher than .6 p-score (experimental group) were copied from the transcripts to Excel spreadsheets and divided into pre-nuclear accented vs. nuclear-accented categories depending on the presence of a boundary with a higher than .6 b-score following them. As a result, the research corpus comprised two sets (RPT 1 and RPT 2), which were compared for similarity with the native-speaker annotations.

Disagreement analysis: As it was mentioned earlier, the essence of the RPT method lies in identifying disagreements among annotators, as they are believed to present the points of perceptual ambiguity driven either by the speech signal itself (a combination of acoustic factors) or the annotators' perception (Cole & Shattuck-Hufnagel 2016). As a result, prominences and boundaries with a high p/b-score (>.6) marked by both the Czech annotators and the native speakers were analyzed in no further detail as they were assumed to be the points of annotators' agreement. Both the experimental and control group members were believed to have reacted to the same/similar acoustic and contextual cues in identifying the prosodic features in question. At the same time, the points of disagreement (the prominences/boundaries marked by one group (>.6) but ignored by the other (<.6)) attracted our special attention. The so-called 'omissions' – prominences and boundaries with a low p/b-score <.6, marked by one group but skipped by the other – were investigated in more detail with the help of additional acoustic analysis (Praat) to identify the factors contributing to the annotation discrepancies.

Acoustic analysis: The words marked by the control group (the native speakers) but not perceived as prominent or preceding a boundary by experimental group annotators and the words marked as prominent by the Czech annotators but unmarked by the native speakers were isolated from the recordings in tone groups (stretches of speech between two boundaries including on average four-to-six phonetic words) and analyzed for (1) the presence/absence of max intensity on the stressed vowel of the word compared to the other words in the tone group, (2) the presence/absence of a tangible pitch change (log max f0) on this word, (3) the presence/absence of a pause (> 0.2 s) following the word, and (4) the word's phone rate, assessed as exceeding/non-exceeding the average phone rate in the tone group, which was calculated as a ratio of the whole tone group duration to the number of syllables in it.

We admit the absence of a universal baseline for the acoustic analysis of all the words subjected to the acoustic analysis connected with the fact that the experimental material involved four different speakers. However, following Styler's (2023: 30) recommendations for using Praat for scientific research,

although different speakers have different baseline values, it is claimed to be acceptable to access maximum/average/minimum values of the inspected features based on a larger than a word excerpt "in comparison with other tokens from the same speaker" if other conditions are statistically normalized.

Statistical analysis: Following Paik (2003), Student's t-test for correlated means (the quantitative measurements of the experimental group's performance in identifying pre-nuclear and nuclear prominences in four texts before and after the training) was additionally conducted to assess the role of the phonological training and annotation practice. It aimed to determine whether the differences in mean numbers of identified prominences in Week 1 (RPT 1, before the training course) and Week 14 (RPT 2, after the phonological training) were statistically significant. MedStatistic online software (https://medstatistic.ru/calculators/calcpars.html) was used to calculate the t-test.

We acknowledge that Student's t-test for correlated means only assesses the overall average ratios of prominences and boundaries in the experimental group's annotations while ignoring the exact locations of the marked features in the transcripts. However, in combination with a non-parametric disagreement analysis, which involved the comparison of every word in the experimental group's transcripts to the native speakers' ground truth, Student's t-test, although with certain reservations, can guarantee the present experiment a desired degree of statistical reliability and declare the presence/absence of the leaning intervention's meaningful effect.

Finally, the experimental group's (RPT 2) annotations were compared with the control groups' data for similarity. Two One-Sided t-tests (TOST) were used to establish the annotations' statistical equivalence (non-inferiority), with the statistical parameters calculated in DATAtab software (https://datatab.net/statistics-calculator/equivalence-non-inferiority), following Lakens et al. (2018). The equivalence test was specifically employed to verify our research hypothesis concerning non-native annotators' ability to perform reliable annotations in RPT.

4 Results

4.1 Stage 1: RPT 1 annotations

The quantitative analysis of the prominences and boundaries identified in the first experimental stage (RPT 1) revealed that without special training non-native annotators are less sensitive to English prominence and boundaries than native speakers of English. In all four recordings, the experimental group marked, on average, 43 per cent fewer prominences than the native-speaker control group (Tables 3, 4, and 6).

	Recording 1	Recording 2	Recording 3	Recording 4
Prominences (raw)	42	45	47	59
Nuclear accented words	38	37	45	55
Pre-nuclear accented words	4	8	2	4
Boundaries (raw)	38	37	45	55

Table 3: The raw number of prominent words and boundaries identified by the experimental group (Czech annotators) in RPT 1

	Recording 1	Recording 2	Recording 3	Recording 4
Prominences (raw)	74	76	96	91
Nuclear accented words	51	44	66	69
Pre-nuclear accented words	23	32	30	22
Boundaries (raw)	51	44	66	69

Table 4: The raw number of prominent words and boundaries identified by the control group (the native speakers)

The total number of spotted prominences in the research corpus was only 57 per cent (mean) of those identified by the control group (the native speakers). The detailed analysis of prominence categories revealed that identifying nuclear prominences, as Bishop (2020) predicted, was less problematic for Czech speakers than pre-nuclear prominence identification (Table 6).

The number of marked nuclear prominences (prominent words preceding a boundary) is, on average, 76 per cent, whereas the number of identified pre-nuclear prominences is only 17 per cent. These figures demonstrate that the Czech annotators' general reliability decreases mainly due to their limited ability to spot pre-nuclear prominences (Table 6).

The relative success of the experimental group in nuclear prominence (76%) and boundary (76%) identification can be explained by the systemic and semantic similarities of boundary tones in English and Czech (e.g. the meaning of finality vs. continuity acoustically expressed at the end of the utterance), which can be easily identified by the Czech speakers even without prior phonological training. At the same time, the experimental group's perceptual challenge in pre-nuclear prominence identification is attributed to frequency and realizational prosodic differences between the two languages, resulting in Czech speakers' inability to identify English prominence without a following boundary.

As described earlier, English primarily relies on the acoustic marking of pre-nuclear accents, such as higher pitch and increased vowel length. However, Czech speakers tend to ignore these down-up acoustic cues, as they do not associate these parameters with stressed syllables in their L1 (Volín et al. 2015, Skarnitzl & Hledíková 2022). Additionally, the narrower pitch range and longer

intonation phrases typical of their mother tongue make pitch peak values typical of English pre-nuclear accents (e.g. H*) irrelevant distractors unless specially trained in phonology classes.

4.2 Stage 2: RPT 2 annotations

The results of the second-stage analysis (RPT 2) demonstrated a tangible increase in the number of identified prominences, significantly shrinking the differences with the native speakers' data (Tables 5 and 6).

	Recording 1	Recording 2	Recording 3	Recording 4
Prominences (raw)	56	62	77	80
Nuclear accented words	48	46	55	61
Pre-nuclear accented words	8	18	22	19
Boundaries (raw)	48	46	55	61

Table 5: The raw number of prominent words and boundaries identified by the experimental group (Czech annotators) in RPT 2

This time, after additional phonological training, the non-native listeners spotted nearly all nuclear prominences (92%) and significantly improved their pre-nuclear perception (62%), resulting in 82 per cent of prominent words (in total) and 92 per cent of boundaries successfully marked.

	RPT 1 (Stage 1)	RPT 2 (Stage 2)
Prominences	57%	82%
Nuclear prominences	76%	92%
Pre-nuclear prominences	17%	62%
Boundaries	76%	92%

Table 6: The statistics of prominent words and boundaries identified by the experimental group before (Stage 1) and after (Stage 2) the additional phonological training

The improved results in prominence and boundary identification (Table 6) can be attributed to the learners' extended exposure to various annotation practices that enhanced their annotation skills, as predicted by McAndrews (2019). On the other hand, the improved ability to spot pre-nuclear accents, which was the most problematic issue in the previous experimental stage (Stage 1, before the additional training), can be explained by the shift, postulated by Schmidt et al. (2016: 245), from the "general auditory sensitivity" to the "contextual perception of suprasegmental information", which manifests itself as experienced listeners' ability to combine down-up and top-down characteristics on the one hand and overcome L1 habits on the other.

4.3 The disagreement analysis

As the native-speaker control group identified significantly more prominences and boundaries in the recordings than the experimental group, there were multiple cases of prominences and boundaries marked by the native speakers but omitted by the Czech annotators (146 individual cases, 43% of all the prominences and boundaries marked by the control group). Yet there were only a few instances of prominent words and boundaries marked by the Czech annotators but unmarked by the native speakers (14 cases, < 1%). After the learning intervention, the number of omitted/added English prominences and boundaries decreased considerably (64 (18%) and 4 (<1%) respective cases).

The analysis of native vs. non-native speakers' perceptual disagreements: The acoustic analysis of 64 words that were marked as prominent by native speakers but were left unmarked by the experimental group even after the learning intervention did not demonstrate consistent acoustic cues in the following Praat analysis (Figure 3). The words were randomly characterized by peak pitch values (29%), increased phone rate (25%), and high intensity (23%), which, however, were not perceived by the experimental group as relevant acoustic cues for prominence marking. The boundary analysis demonstrated that the boundaries perceived as such by the native speakers but unmarked by the Czech group were all shorter than .2 seconds, which classifies them as 'breath pauses' rather than 'syntactic pauses', typically considered more tangible for non-native listeners (Igras-Cybulska et al. 2016).

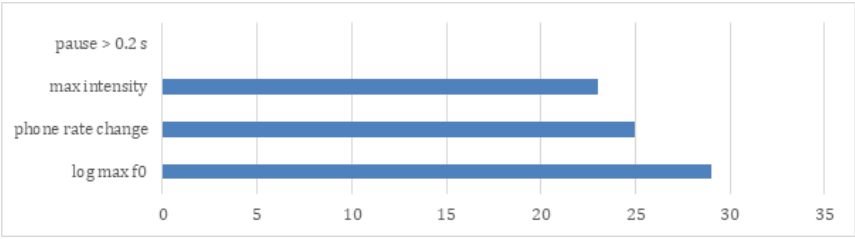

Figure 3: Acoustic characteristics of prominences and boundaries omitted by the Czech annotators but marked by the native speakers

Neither did the acoustic analysis of 14 words marked by the Czech annotators as prominent and/or preceding a boundary demonstrate a consistent acoustic coloring, which could have contributed to their false recognition (Figure 4). The only thing worth noting in their acoustic properties is that 40 per cent of

them were followed by a longer than a .2 second pause, possibly entailing their marking by the Czech annotators (Example 4).

(4) | *for just four days*| ***so*** | *my first four days were critical* | – The word "so", followed by a 0.23 second breath pause was falsely marked as prominent and preceding a boundary although neither the prominence nor the boundary were marked by the native speakers in this case.

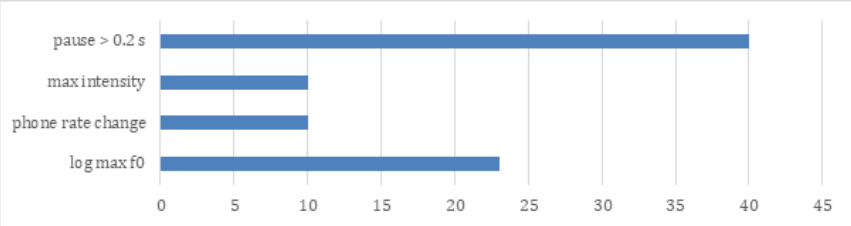

Figure 4: Acoustic characteristics of prominences and boundaries marked by the Czech annotators but unmarked by the native speakers

The analysis of RPT 1 vs. RPT 2 perceptual disagreements: The prominences/boundaries left out by the Czech annotators in Stage 1 but marked in Stage 2 attracted our special attention (Example 5).

(5) |*I cannot* ***remember*** *telling my* ***parents***| – The pre-nuclear accent "remember" in this tone group was left unmarked by the experimental group in RPT 1. Still, it was marked as prominent in the RPT 2 stage. The Praat analysis demonstrated that the word is not colored by a peak intensity or a noticeable change in the phone rate but carries a distinct peak pitch accent (H*), recognized by the experimental group after the learning intervention.

The computer-assisted analysis of the main acoustic characteristics of 82 words that were not identified as prominent in the first experimental stage (RPT 1) but were marked as prominent in Stage 2 (RPT 2) demonstrated that some acoustic characteristics of the omitted words were more common than the others (Figure 5). Eighty-three per cent of the words carried a peak pitch value (log max f0) in the respective tone group, 32 per cent had an extended phone rate (an increased length per syllable), and 18 per cent were characterized by increased loudness (a higher-than-average intensity in the tone group). Only less than five per cent were followed by a pause longer than .2 seconds.

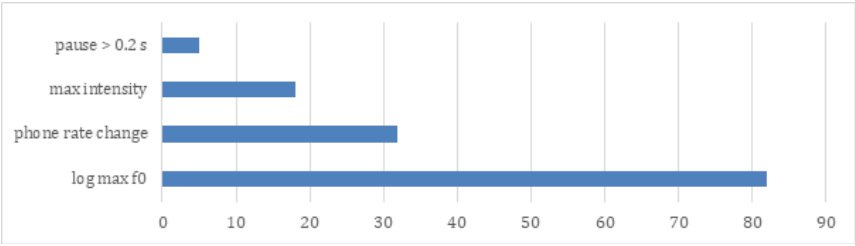

Figure 5: Acoustic characteristics of prominences and boundaries unnoticed by the Czech annotators in RPT 1 but marked by them in RPT 2

It can be concluded that the most meaningful acoustic parameter, whose perception was significantly improved after the additional phonological training course, was the peak pitch value (e.g. H* accents), regarded as the basic cue for English pre-nuclear prominence perception. The improved statistics of its identification after the learning intervention demonstrates that the training course led to a notable improvement in the Czech annotators' perception of overall prominence and boundary.

4.4 Statistical analysis

The results of Student's t-test for correlated means, which was conducted to determine if the difference between the number of prominences and boundaries marked by the experimental group in Week 1 (before the training course) and in Week 14 (after the training) was meaningful, confirmed that the change in the number of identified prominences and boundaries was statistically significant ($t(11)=3.73$, $p=.005$). The obtained t-value proves the impact of additional phonological training on non-native speakers' perception of English prominence.

The TOST equivalence tests confirmed the statistical equivalence of the experimental group's (RPT 2) and the control group's RTP annotations (with 90% desired power). The graphical representation of the TOST measurements (Figure 6) demonstrates that the mean difference in both groups' RPT annotations lies within the established confidence interval of min./max. SD=±10.

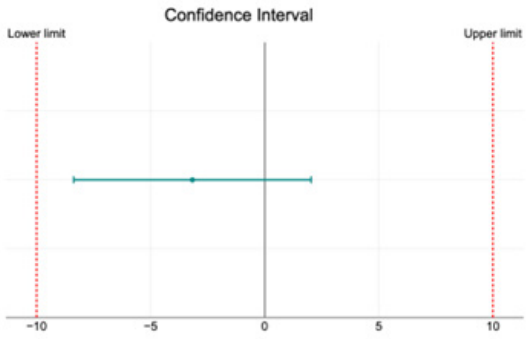

Figure 6: The results of the equivalence testing presented on a confidence interval (-10, 10) scale

The equivalence of the two RPT data sets is also illustrated by the groups' boxplots (Figure 7). The boxplots show the distributions of the collected means of prominences and boundaries in four monologues as identified by the experimental vs. control groups. A visible horizontal intersection of the boxplots standing for the respective data sets shows a minimal discrepancy between the quantitative measurements of non-native vs. native annotations in our experiment.

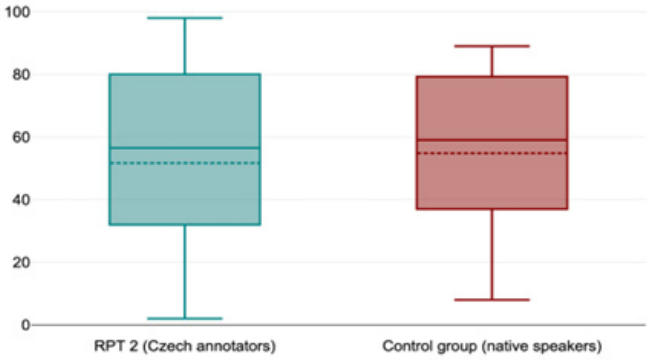

Figure 7: The results of equivalence testing presented as boxplots

As a limitation of the conducted statistical analysis, it is necessary to acknowledge that both parametric tests (Student's t-test and the TOST equivalence

test) compared only the raw counts of prominences and boundaries identified by the Czech annotators against the native speakers' ground truth, while the exact locations in the texts remained statistically unassessed and need further analysis.

5 Discussion

The results of the auditory analysis performed by Czech learners of English (non-native listeners) compared with the auditory data from native English speakers demonstrate that in the tested prosodic parameters, namely boundary and prominence identification, language-proficient Czech speakers of English without additional suprasegmental training demonstrate weaker sensitivity to English prominence and boundaries than native speakers of English. This difference is especially noticeable in marking pre-nuclear prominence, which is believed to be more challenging to perceive due to the existing realizational and frequency differences between Czech and English phonological systems.

The computer-assisted acoustic analysis of the words that were marked as prominent by native speakers of English but skipped by phonologically naïve Czech listeners (the disagreement analysis) demonstrated that of all the parameters measured, changes in f0 were the most likely problematic cues, whose targeted training during a fourteen-week learning intervention led to an improvement in prominence/boundary marking in the post-intervention RPT stage. As a result, after the training, phonologically trained Czech speakers of English provided more reliable prosodic annotations that were statistically equivalent to the British native speakers' markings.

The comparison of correlated means (Czech learners' prominence and boundary marking of four semi-spontaneous English monologues in RPT 1 and RPT 2) proves that, quite as predicted by earlier studies into L2 learners' suprasegmental perception (Schmidt et al. 2016, van Maastricht et al. 2017, McAndrews 2019, Tymbay 2024), the learning intervention improved the perception of key prosodic parameters, including the problematic pre-nuclear prominence. We believe that the suggested training course increased the learners' targeted exposure to the suprasegmental elements of English prosody and helped them, to a certain degree, overcome the L1 transfer stemming from the phonological differences between English and Czech.

The low sensitivity of Czech (L1) listeners to English pre-nuclear acoustic signals partly supports the mentioned Markedness Differential Hypothesis (Eckman 1977) and follow-up research (Mennen 2015) into cross-language speech perception, which draw learners' attention to the marked prosodic features of the foreign language as crucial yet most challenging to master, especially if they are not marked in the student's L1.

The review of literature shows that typical of the Czech language, insufficient attention to the acoustic marking of prominence, illustrated, for instance, by the absence of the stressed vowel lengthening and narrower pitch range, may interfere with English pre-nuclear prominence perception. Since pre-nuclear prominence in English is primarily signal-driven, certain aspects of Czech-to-English prosodic interference, such as various degrees of stress and accent, the inventory of pitch accents and positional vowel changes, should be further highlighted during preparatory phonology classes to improve Czech students' English prosody perception. This way, additional identification, comprehension, and discrimination practice supplemented by various annotation exercises will assist in addressing the challenge.

6 Conclusions

The experimental study into Czech speakers' perception of crucial English suprasegmental features was conducted through longitudinal classroom research, including an intensive fourteen-week-long suprasegmental training and two-stage gauging of learners' ability to detect English prominence and boundaries. The study's results prove that phonologically-trained English-proficient Czech speakers possess a sustained ability to adequately perceive English prominence and boundaries and mark them using the RPT convention.

This paper explains the instances of low sensitivity of phonologically naïve listeners to specific suprasegmental parameters, such as English pre-nuclear prominence, by the negative prosodic transfer of the learners' native language. Therefore, introducing contrastive language practice into foreign language teaching may contribute to a more consistent specification of the target language's marked features and build better awareness of them, especially if these characteristics are realized differently or present less frequently in the student's mother tongue.

Our study demonstrates that recognition of English prominence among L2 students can be significantly improved by integrating various perceptual exercises directly into English teaching. Meaningful prominence identification assisted by RPT practice provides English learners with valuable experience that they can rely on to acquire better differentiation and comprehension skills.

Involving L2 learners in intonation studies is also crucial for the potential validation of English prosodic research in non-English speaking countries since the availability of phonologically trained native speakers familiar with a specific pronunciation variety may be a challenge. While validation by native speakers is valuable in many cases, it is essential to note that non-native listeners can still contribute to phonetic studies, especially when their expertise and training in

phonetics are robust. Collaborations between native and non-native researchers seem to be another fruitful way to ensure accuracy and broader insights into English phonetic studies conducted in non-English speaking countries.

Finally, as a limitation of the present study, we would note insufficient attention paid to individual annotators' reports, as the study was primarily focused on the Czech annotators' mid-sized group performance. A more detailed statistical procedure involving some non-parametric methods is also needed to consider the delicate nuances of prominence/boundary placement and the annotators' individual work.

Appendix

The main components of the SPELB intervention course

Stress patterns: First, the SPELB course addressed the Czech students' inconsistent stress placement and issues with pronouncing unstressed syllables in English. We developed a multi-stage training program focusing on stress patterns different from the learner's L1 (Czech) (Example 1).

1. Some English words have a double stress, which is impossible in Czech. As a rule, these are polysyllabic or compound words. The first syllable usually carries the primary stress, with the secondary stress falling on the latter syllables. Read the words observing the primary and secondary stress:

   handbag – /ˈhænd ˌbæg/
   handicap – / ˈhændɪ ˌkæp/
   Amsterdam – /ˈæmstə ˌdæm/
   haberdashery – /ˌhæbəˈdæʃərɪ/

By conducting smartphone-assisted recording sessions, the students regularly self-assessed themselves and their peers, tracking their progress to ensure they consistently used correct stress placement in their speech. By comparing and contrasting the recorded utterances to the native speakers' recordings, the SPELB course additionally accentuated the role of various acoustic cues in achieving speech intelligibility.

Unstressed vowel practice: We implemented targeted training for Czech speakers on pronouncing unstressed syllables, mainly when dealing with reduced vowel 'schwa' (/ə/).

Since vowels in unstressed syllables of polysyllabic words and weak grammatical word forms are mostly reduced and centralized toward the mid-central vowel schwa (/ə/), learners were trained to recognize and reproduce

this centralization. The course provided practice exercises focusing on differentiating schwa from other vowel sounds (Example 2).

2. Read the tongue twister, but first identify in which words the vowel "a" is reduced to the schwa /ə/ sound and where it is pronounced as the /æ/ sound.

   Can you can a can as a canner can can a can?

Melody and pitch range: To address the challenges of limited pitch range and melodic patterning, we encouraged speakers to diversify their pitch range by stressing the phonological differences of different tones and the critical role they play in conveying emotions and attitudes in English speech. It was achieved through focused imitation exercises and vocal drills to expand pitch variety (Example 3).

3. Read the following sentences using different types of nuclear tones. First, practice Low Falls on the words in bold, then High Falls, Low Rises, Fall-Rises, and Mid-Levels. Try to assess the differences in the meaning of the phrase said with different intonation patterns:

- 1. ***Darling***!
- 2. I can't see our ***car***!
- 3. Where did you ***park*** it?

We integrated specific exercises targeting melodic steps between stressed and unstressed syllables, encompassing traditional British annotations – falling tones vs. rising tones – and ToBI elements – H* vs. L*. The latter approach contributed to better segmentation of intonation patterns, ultimately enhancing overall intonation recognition.

Additionally, to help students better accommodate English intonation patterns, we used audio recordings of native speakers as models for varied pitch patterns. We encouraged Czech speakers to mimic and practice these patterns in different contexts (Example 4).

4. Listen to and read the limerick, paying particular attention to the intonation:

   Whether the weather is (↗) fine or whether the weather is (↘) not.…whether we (↗) like it or (↘) not.

Note to the exercise: This English limerick has the intonation of enumeration: rising tones on all the objects (or their characteristics) in the list and a fall on the last word.

English rhythm: Understanding the unique aspects of rhythmic patterning in English compared to Czech was essential for our learners as these distinctions impact spoken English's natural flow and rhythm.

Practical exercises modelling native speakers' text-integrated full/reduced vowel alteration were incredibly beneficial, as this practice contributed to a more natural and fluid English speech rhythm (Example 5).

5. Identify the words with the /ə/ sound. Scan the QR code and listen to the dialogue, then read it, paying attention to the alteration of stressed and unstressed vowels.

   - Paul! All my in-laws are coming in August! They're forcing us to board them.
   - It's awful. You ought to call them! We can't afford it!
   - I've already called! They say our house is gorgeous, but I think it's rather small!

The course also emphasized the significance of connected speech processes facilitating English rhythm. We engaged learners with practical examples and exercises to hone these skills and understand their impact on the flow of English speech. To enhance English rhythmic patterning, we focused on improving the utilization of linking and glottalization in speech, which are especially problematic for Czech (L1) speakers (Example 6).

6. Note that, although usually silent in combinations with vowels (British English), the letter "r" becomes pronounced as the sound /r/ if the word following it begins with a vowel sound. This position is called the linking "r". Read the exercise, paying attention to the linking "r" sound.

   You are ill.
   They are in.
   I can't hear anything.

We provided regular feedback and corrections to ensure learners achieved accurate and natural pronunciation. By following these recommendations, our learners better understood the rhythmic distinctions between English and Czech and improved their perception of natural English rhythm.

References

Adams, C. and Munro, R. (1978) 'In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English.' *Phonetica 35*(3), 125-156. https://doi.org/10.1159/000259926

Baese-Berk, M., McLaughlin, D. and McGowan, K. (2020) 'Perception of non-native speech.' *Language and Linguistics Compass 14*(7), e12375. https://doi.org/10.1111/lnc3.12375

Bishop, J., Kuo, G. and Boram, K. (2020) 'Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription.' *Journal of Phonetics 82*, 100977. 1-20. https://doi.org/10.1016/j.wocn.2020.100977

Brière, E. J. (1966) 'An investigation of phonological interference.' *Language 42*(4), 768-796.

Chamonikolasová, J. (2018) *Intonation in English and Czech Dialogues.* Brno: Masarykova univerzita.

Cho, T. and Keating, P. (2009) 'Effects of initial position versus prominence in English.' *Journal of Phonetics 37*(4), 466-485. https://doi.org/10.1016/j.wocn.2009.08.001

Čtvrtečková, L., Fischer, O. and Skarnitzl, R. (2023) 'Pronunciation of Czech secondary school English teachers and their cognition about pronunciation.' In: Henderson, A. and Kirkova-Naskova, A. (eds) *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices*. 31-45. https://doi.org/10.5281/zenodo.817403

Cole, J., Mahrt, T. and Roy, J. (2017) 'Crowdsourcing prosodic annotation.' *Computer Speech & Language 45*, 300-325. https://doi.org/10.1016/j.csl.2017.02.002

Cole, J. and Shattuck-Hufnagel, S. (2016) 'New methods for prosodic transcription: Capturing variability as a source of information.' *Laboratory Phonology 7*(1), 1-29. https://doi.org/10.5334/labphon.29

DATA tab software (2024) Available online at https://datatab.net/statistics-calculator/equivalence-non-inferiority

Demolin, D. (2012) 'Experimental methods in phonology.' TIPA. *Travaux Interdisciplinaires sur la Parole et le Langage 28* (online). http://doi.org/10.4000/tipa.162

Eckman, F. (1977) 'Markedness and the contrastive analysis hypothesis.' *Language Learning 27*, 315-330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Flege, J. E., Schirru, C. and MacKay, I. R. (2003) 'Interaction between the native and second language phonetic subsystems.' *Speech Communication 40*(4), 467-491. https://doi.org/10.1016/S0167-6393(02)00128-0

Foulkes, P. and Docherty, G. J. (2006) 'The social life of phonetics and phonology.' *Journal of Phonetics 34*, 409-438. https://doi.org/10.1016/j.wocn.2005.08.002

Grice, M. and Kügler, F. (2021) 'Prosodic prominence – a crosslinguistic perspective.' *Language and Speech 64*(2), 253-260. https://doi.org/10.1177/00238309211015768

Hume, E. (2004) 'Markedness: A predictability-based approach.' *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 30*(1), 182-198. https://doi.org/10.3765/bls.v30i1.948

Jenkins, J. (1995) *Variation in Phonological Error in Interlanguage Talk.* PhD dissertation. London: Institute of Education, University of London.

Igras-Cybulska, M., Ziółko, B. and Żelasko, P. (2016) 'Structure of pauses in speech in the context of speaker verification and classification of speech type.' EURASIP *Journal on Audio, Speech, and Music Processing 18*, 1-16. https://doi.org/10.1186/s13636-016-0096-7

Kirkova-Naskova, A. (2010) 'Native speaker perceptions of accented speech: The English pronunciation of Macedonian EFL learners.' *Research in Language 8*, 41-61. https://doi.org/10.2478/v10015-010-0004-7

Lakens, D., Scheel, A. M. and Isager, P. M. (2018) 'Equivalence testing for psychological research: A tutorial.' *Advances in Methods and Practices in Psychological Science 1*(2), 259-269. https://doi.org/10.1177/2515245918770963

Lee, B., Plonsky, L. and Saito, K. (2020) 'The effects of perception- vs. production-based pronunciation instruction.' *System 88*, 102185. https://doi.org/10.1016/j.system.2019.102185

Mahrt, T. (2016) 'LMEDS: A platform for collecting prosodic annotations online.' In: *Forschungskolloquium SS.* Cologne, Germany. Online document. Retrived on 24 September 2024. https://hal.science/hal-01469868

MedStatistic online software (2013) Available online at https://medstatistic.ru/calculators/calcpars.html

Mennen, I. (2015) 'Beyond segments: Towards an L2 intonation learning theory.' In: Delais-Roussarie, E., Avanzi, M. and Herment, S. (eds) *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations.* Berlin: Springer. 171-188. https://doi.org/10.1007/978-3-662-45168-7_9

McAndrews, M. (2019) 'Short periods of instruction improve learners' phonological categories for L2 suprasegmental features.' *System 82*, 151-160. https://doi.org/10.1016/j.system.2019.04.007

Paik, C. (2003) *Statistical Methods for Rates and Proportions.* New Jersey: John Wiley and Sons.

Prat 6.4.03 (2024) Available online at https://www.fon.hum.uva.nl/praat

Schmidt, E., Kung, C., Post, B., Yuen, I. and Demuth, K. (2016) 'L1 experience shapes the perception of intonational contours.' *Proceedings of the Intertional Conference on Speech Prosody, Boston.* 242-246.

Skarnitzl, R. and Hledíková, H. (2022) 'Prosodic phrasing of good speakers in English and Czech.' *Frontiers in Psychology 13*, 857647. https://doi.org/10.3389/fpsyg.2022.857647

Skarnitzl, R. and Rumlová, J. (2019) 'Phonetic aspects of strongly-accented Czech speakers of English.' *AUC Philologica 2019*(2), 109-128. https://doi.org/10.14712/24646830.2019.21

Smith, C. L. and Edmunds, P. (2013) 'Native English listeners' perceptions of prosody in L1 and L2 reading.' *Proceedings of the Annual Conference of the International Speech Communication Association*, *INTERSPEECH 2013*. 235-238. https://doi:10.21437/Interspeech.2013-74

Spreafico, L. and Vietti, A. (2022) 'Techniques and methods for investigating speech articulation: The centrality of instruments.' *Laboratory Phonology 13*(1), 1-8. https://doi.org/10.16995/labphon.8538

Styler, W. (2023) *Using Praat for Linguistic Research.* University of Colorado at Boulder Phonetics Laboratory. Available online at https://wstyler.ucsd.edu/praat/

Tremblay, K. L., Piskosz, M. and Souza, P. (2003) 'Effects of age and age-related hearing loss on the neural representation of speech cues.' *Clinical neurophysiology 114*(7), 1332-1343.

Trofimovich, P. and Baker, W. (2006) 'Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech.' *Studies in Second Language Acquisition 28*(1), 1-30. https://doi.org/10.1017/S0272263106060013

Trubetzkoy, N. (1939) *Grundzüge der Phonologie.* Göttingen: van der Hoeck and Ruprecht.

Tymbay, A. (2022) 'Prominence and melody in the spontaneous speech of Czech learners of English.' *SKASE Journal of Theoretical Linguistics 19*(2), 2-17.

Tymbay, A. (2024) 'On non-native listeners' ability to identify prominence and pitch accents in English monologic speech.' *ELOPE English Language Overseas Perspectives and Enquiries 21*(1), 63-88. https://doi.org/10.4312/elope.21.1.63-88

van Maastricht, L., Zee, T., Krahmer, E. and Swerts, M. (2017) 'L1 perceptions of L2 prosody: The interplay between intonation, rhythm, and speech rate and their contribution to accentedness and comprehensibility.' *INTERSPEECH 2017*-1150, 364-368.

Volín, J., Poesová, K. and Weingartová, L. (2015) 'Speech melody properties in English, Czech and Czech English: Reference and interference.' *Research in Language 13*(1), 107-123. https://doi.org/10.1515/rela-2015-0018

Volín, J. and Zimmermann, J. (2011) 'Spectral slope parameters and detection of word stress.' In: Byron, P. (ed.) *Technical Computing Prague 2011*, 125-129. Praha: Humsoft. Online document. Retrived on 24 September 2024 <https://dsp.vscht.cz/konference_matlab/MATLAB11/prispevky/125_volin.pdf>.

Weingartová, L., Poesová, K. and Volín, J. (2014) 'Prominence contrasts in Czech English as a predictor of learner's proficiency.' In: *Proceedings of the International Conference on Speech Prosody 2014.* 236-240. https://doi.org/10.21437/SpeechProsody.2014-35

Alexey Tymbay is Assistant Professor at the Department of English, Technical University of Liberec, and a member of the Czech Association for the Study of English (CZASE). He teaches multiple linguistic courses, including EFL, ESP, and other subjects. He has authored several papers on English phonology, political discourse analysis, and communicative linguistics.
ORCID ID: https://orcid.org/0000-0003-2061-0263

Address: Alexey Tymbay, Department of English, Faculty of Science, Humanities and Education, Technical University of Liberec, Komenského 314, Liberec 46005, Czech Republic. [e-mail: alexey.tymbay@tul.cz]