



***StudyIntonation* courseware kit for EFL prosody teaching**

*Elena Boitsova*¹, *Evgeny Pyshkin*², *Yasuta Takako*³, *Natalia Bogach*⁴,
*Iurii Lezhenin*⁴, *Anton Lamtev*⁴, *Vadim Diachkov*⁴

¹Institute of Humanities Peter the Great St. Petersburg Polytechnic University, Russia

²Software Engineering Lab. University of Aizu, Japan

³Center for Language Research. University of Aizu, Japan

⁴Institute of Computer Science and Technology Peter the Great St. Petersburg Polytechnic University, Russia

el-boitsova@yandex.ru, pyshe@u-aizu.ac.jp, yasuta@u-aizu.ac.jp,
bogach@kspt.icc.spbstu.ru, lezhenin@kspt.icc.spbstu.ru, lamtev@kspt.icc.spbstu.ru,
dyachkov@kspt.icc.spbstu.ru

Abstract

The paper investigates EFL prosody teaching and learning obstacles in monolingual societies alongside the main limitations for efficient deployment of specific prosody training tools based on audiovisual feedback. Prosody teaching and learning environment *StudyIntonation* containing the mobile application (MA) and the courseware development kit (CDK) is examined from teacher perspective. In *StudyIntonation* approach teachers are supposed to be co-authors of learning content and can contribute to new courses development. *StudyIntonation* approach puts together the advances of pitch visualization technology with Android mobile application development to offer a handy open-source prosody teaching, learning and research platform. It is shown, that new courses may be launched within a very short period provided native designed or spontaneous speech records are available.

Index Terms: CALL, human-computer interaction, prosody teaching.

1. Introduction

Prosody is a very complex phenomenon which involves many aspects of speech, including stress, rhythm, and intonation, collectively called suprasegmentals [1]. The importance of prosody teaching is supported by extensive research and has been repeatedly highlighted in numerous papers. Though there is no universal way to access L2 speech intelligibility, it is acknowledged, that poor prosody influences it negatively rather than phonetic errors or accent (e. g. see [2]). Nevertheless, EFL prosody training is often marginalized and remains one of the most problematic points in modern teaching practice due to a number of reasons:

Firstly, learners are generally reluctant to prosody training and find specific pronunciation exercises tedious, time-consuming, discouraging and boring [3], [4]. The authors' personal experience shows that in monolingual societies like Japan or Russia this situation is even worse, because, without everyday need, students have psychological barriers in speaking. Being involved in pronunciation trainings, many of them try to ignore or escape the offered activities.

Secondly, teachers do not pay much attention to prosody exercises. This could be explained either by lack of time in academic curricula or by learning priorities other than pronunciation training. There is also multiple evidence of generally insufficient teachers' preparation in the area of pronunciation

instruction. D. Chun in [1] and M. Derwing et al. in [2] provide a critical analysis of the current state.

Thirdly, there are excellent projects of specific prosody teaching software (e. g. [5], [6] [7], [8], [9]), which yet being a great help can be further improved [10], [11], [12]. Since displays of pitch contours help both perception and production of tone and intonation (e. g. see [13]) this software includes usually an automatic speech recognizer (ASR), pitch processing tool (e. g. *Praat*) to produce a visual acoustic and/or articulatory display of a native-speaker model as well as a number of other multimedia facilities. A comprehensive survey on ASR-based tools with audiovisual feedback can be found in [14].

On the one hand, these tools and programs often lack teacher and learner instructions on how to interpret the difference between model and speaker; on the other hand, since teachers are not involved in courseware development, the majority of training software lack pedagogically sound content [15].

To overcome the first issue learners need training on the visualizations of their pitch contours to improve their pronunciation. These training needs to be implemented as software built-in modules but are not currently available in an absolute majority of teaching software[1].

The second issue can be at least partially solved by offering ways in which classroom teachers as researchers can look for answers on their own [16] in case the instructors could collaborate with linguists and computer scientists to design pedagogically valuable exercises thus bridging the gap between scientific findings, multimedia programming and everyday teaching practice [15], [2], [1]. This implies that teachers will not be merely passive software users, but will gain a position of equal participants of teaching software design. Levis in [15] claims that teachers should have the opportunity to develop and test their own computer-based pronunciation exercises through the use of CALL tools.

This research examines *StudyIntonation*¹ [17] exactly from teacher perspective and claims its availability for teachers to participate in courseware preparation. *StudyIntonation* has no obstacles to create prosodic activities or visual add-ons to the existing pronunciation courses. *StudyIntonation* Courseware content is accessed via mobile application aimed at giving EFL students a digital gaming environment [18], making the learning process less stressful and more exiting because they can shift from classroom work to mobile interaction.

¹<https://bogach.github.io/prof-higgins/>

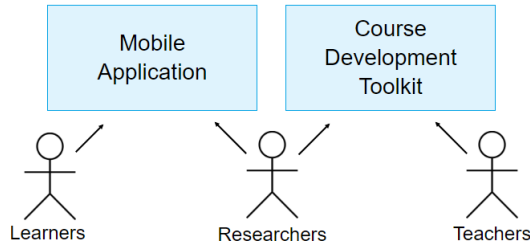


Figure 1: Target groups

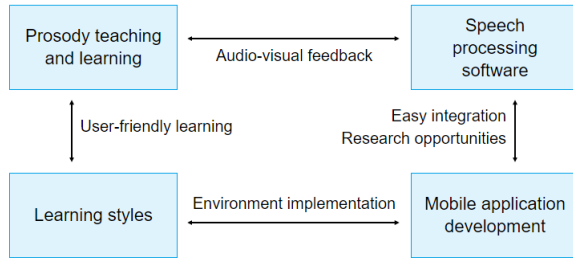


Figure 2: Study Intonation Approach

The rest of the paper is organized as follows: Section 2 gives a brief description of *StudyIntonation* approach, Section 3 discloses the steps of courseware preparing to support the thesis that, using *StudyIntonation*, teachers are able to create their own courses on any topics and of any level of complexity, Section 4 provides discussion on prosody teaching problems in monolingual societies and analyses *StudyIntonation* benefits as a CALL tool for prosody teaching.

2. Approach and methodology

StudyIntonation approach connects 4 basic ideas (see Fig. 2) [17]:

- Prosody teaching and learning via audiovisual feedback
- Learning styles of students
- Speech processing
- Mobile application development

The learning content organization of *StudyIntonation* addresses the whole diversity of learning styles [17]. The tasks are offered within four typical speech situations. Actually at present the available course has 78 tasks of progressing difficulty explaining a number of phrasal intonation patterns visualized by pitch plots. Pitch estimation based on *TarsosDSP* [19] is used to provide learners with audiovisual feedback. While the use of visualization technology is a crucial advance in teaching, such teaching can be further enhanced by connecting specific pitch processing technology to mobile application development.

StudyIntonation is an open-source extendable CALL environment for Android platform (Fig. 3, 4). Mobile devices with their graphics and sound processing capabilities are widely used for language learning. Nevertheless, mobile language learning products do not focus typically on prosody and pronunciation. The goal to create *StudyIntonation* was to provide learners with a handy, natural, intuitive and friendly application for prosody and pronunciation (Fig. 5). At the same time, with *Study Intonation* teachers can develop series of courses from the most elementary level to the most advanced one. At this point, there

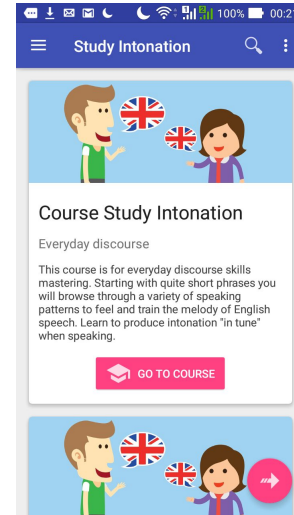


Figure 3: Study Intonation Course page

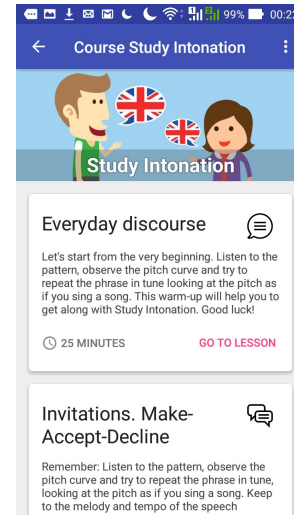


Figure 4: Study Intonation Lesson page

is only one course in *Study Intonation*, however, basically, there is no limit to their number.

StudyIntonation is meant to contribute to the solution of 4 global issues dealing with prosody teaching and learning, especially in monolingual societies [1], [4]:

- EFL prosody training is marginalized and neglected in spite of the fact that prosodic (suprasegmental) errors are more likely to be an obstacle for intelligible speech than phonetic (segmental) ones or accentedness.
- Learners are "shy" and are not eager to practice speaking
- Teachers do not have a specific training to provide efficient speaking instructions concerning suprasegmentals
- CALL software lacks linguistically and pedagogically sound content

Thus, *StudyIntonation* comprising the mobile application (MA) and the courseware development kit (CDK) (Fig. 6) is addressed to 3 target groups: learners, researchers and teachers (Fig. 1). Using MA, students overcome a barrier to speak

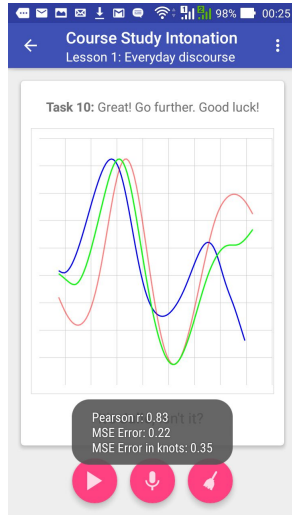


Figure 5: Example of an interrogative question "It's Julie, isn't?" (Lesson 1, Task 10) completed by a student. Model is given in blue. Pearson correlation $r = 0.83$, $MSE = 0.22$.

and become encouraged to train and improve their speaking skills. MA gives EFL students a digital gaming environment and makes the learning process less stressful because MA allows students to switch between mobile interaction and ordinary classwork. To compare model and record curves two quality metrics, r and MSE , are applied jointly:

$$r = \frac{\sum_{i=1}^k |(S^m(\tau_i) - \overline{S^m(\tau)})(S^r(\tau_i) - \overline{S^r(\tau)})|}{\sqrt{\sum_{i=1}^k (S^m(\tau_i) - \overline{S^m(\tau)})^2} \sqrt{\sum_{i=1}^k (S^r(\tau_i) - \overline{S^r(\tau)})^2}}, \quad (1)$$

$$MSE = \frac{1}{k} \sum_{i=1}^k (S^m(\tau_i) - S^r(\tau_i))^2. \quad (2)$$

$S^r(t)$ and $S^m(t)$ are record and model pitch series respectively, τ – time parameter. Result is displayed on the screen and can be exported to a database. For any study on suprasegmental production, *StudyIntonation* provides a tool to develop the original experimental content, export metrics and time series for data analysis.

From teacher perspective, for proficient prosody instructors, *StudyIntonation* provides an open-source platform to create audiovisual training exercises. They can share this advanced knowledge with others, thus, boosting knowledge transfer in a complex and problematic area of prosody training.

3. Courseware design

Courseware design from scratch does not require any programming skills and can be entirely performed by a content-responsible team (teachers, dictors, actors). The whole process involves the following steps (Fig. 6):

1. Record either intendedly pronounced phrases or excerpt passages from extemporaneous speech and store them as .mp3 files. For this purpose, any sound editor (e. g. Audacity) will perfectly do.
2. Provide a text file with marked up text subtitles in the form of $t_{start}, t_{stop}, WORD$ e.g. 2678 2687 "Hello".

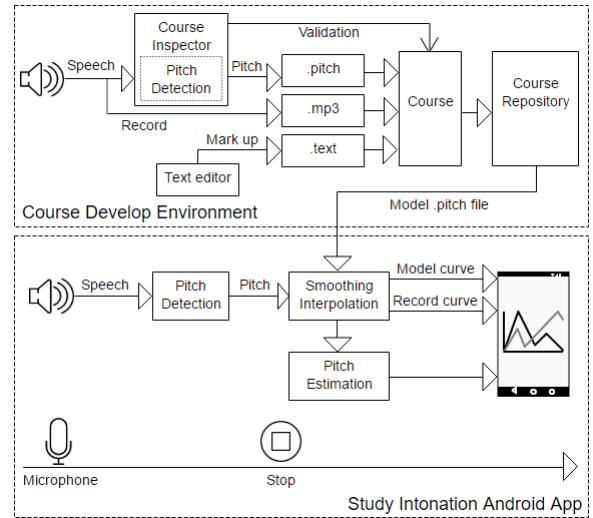


Figure 6: StudyIntonation workflow

3. Input an .mp3 file to *CourseInspector* utility through command line interface. This utility will produce a .pitch file with corresponding pitch sample set. A tuple of .mp3, .pitch and .text files correspond to one task in a lesson within a course. *CourseInspector* also performs the final validation of each task.

Course preparation stage prior to making .mp3 records is implicit. It assumes course content planning, selection and sorting. This operation can be rather long, but it is exactly here, where a teacher is involved and contributes to courseware design. After this stage is fulfilled, launching an updated version of *StudyIntonation* containing a new course can be completed.

Study Intonation could be applied in teaching not only English but whatever foreign language. It can be especially useful for studying languages where the stress is mainly phrasal (here, a typical example is French language) or where the stress is tonical, not dynamic, so different pitches mean different words (this is typical for most of East Asian languages, especially Chinese).

4. Discussion

The deployment scope of CALL tools for prosody teaching in high school of Russia or Japan is quite moderate at present. It can be explained by the fact that Japan and Russia are basically monolingual societies where people can live pretty good without everyday need to communicate in foreign languages. The English education in these countries is focusing on learning grammar rules and memorizing words and expressions necessary for high school/college entrance examinations.

Under such a circumstance, speaking skills are rarely taught in English class, and teaching pronunciation has been almost neglected at both segmental and suprasegmental levels. Teaching pronunciation skills, especially intonation patterns, is one of the least emphasized aspects of English education in both countries.

Typical Japanese English learners transfer their L1 accentuation patterns to English, and produce "flat" English both at the word and sentence levels. Japanese is a pitch accent language which distinguishes high and low in pitch without increasing loudness and length of the stressed syllable. Some dialects in-

cluding Fukushima dialect do not have pitch patterns and are categorized as accentless dialects that have flat word-level pitch contours. The speakers of accentless dialects may especially struggle with accentuation patterns and intonation in English. Russian has seven main pitch patterns [20], that differ greatly from those existing in English. Thus, L1 accentuation transfer of Russian speaking students produces strong accent, yet not always unintelligible.

Basically, in Japan and Russia teachers rarely spend time on actual pronunciation practice. Students are mostly encouraged to memorize accentuation rules and work on written pronunciation drills for entrance examinations, but they may not know how to actually perform what they have memorized. Most Japanese teachers do not teach intonation patterns except for basic rules such as "rising intonation" for yes-no questions.

The second problem is lack of time. Especially in junior high school and high school, teachers must cover certain contents for entrance examinations. As speaking skills are not usually tested in exams, most teachers lessen pronunciation practice.

The third problem is the students learning attitude. Typical Japanese students remain passive and quiet in class, and they are reluctant to participate in speaking activities. Teachers may wish to encourage their students to speak up, but hesitate to teach pronunciation. This is because they are afraid of discouraging their "shy" students by pointing out and correcting their pronunciation errors.

Under these challenges CALL software obviously allows advancing prosody teaching and learning process in several directions:

- Performance and efficiency (i.e. reducing time for knowledge and skills delivery, supporting long-lasting knowledge and skills, as well as easier focus on certain topics),
- Accessibility (enabling personalized access to new materials and collaborative instruments which are hardly available without computers),
- Flexibility (access to learning tools in any time and from any location);

Courses within *Study Intonation* were prepared in a relatively short time. The first stage of courseware design when the appropriate teaching content is selected, sorted and recorded can be rather time-consuming. It assumes a large teacher creativity as well as involvement of native-speakers. After this stage is completed, further steps can be performed very quickly and new courseware may be launched and accessed via MA. Table 1 contains the approximate spent time values for actual *Study Intonation* courses. Approximate time spent for course preparation after content selection stage is shown in Table 1.

Table 1: Time spent for course preparation

| Course module, Tasks | Time, hours |
|----------------------|-------------|
| Everyday phrases/20 | ≈ 12 |
| Invitations/25 | ≈ 36 |
| Emotions/15 | ≈ 12 |
| Conference/15 | ≈ 10 |

5. Conclusion

Converting and/or representing study materials in digital format and automation of original non-digital process is not a unique CALL outcome: rapid technology transformation requires further use of digital technologies and data in order to transform the processes and create an environment where digital technology creates completely new possibilities for language learners and teachers. CALL is not limited to mere digitalization of learning process by transferring traditional techniques of managing language-related data with the use of computers. It creates totally new use cases that are aimed to overcome formerly infeasible challenges.

Study Intonation gives researchers and teachers a quick tool to put annotated content (phrases, dialogues, talks, etc.) together with an audiovisual support and feedback into a form of lessons or experiment tasks. This research examines *Study Intonation* courseware development kit from the teacher perspective. We show that *Study Intonation* approach has no limitations for new courses design and teacher can play a leading role in this process, thus, creating pedagogically valuable teaching content.

Modified accordingly, *Study Intonation* may be applied as a research tool in phonetics, phonology and communicative linguistics and, also, in the experimental phonetic studies of dialects.

Other possible applications of *Study Intonation* could be those regarding intonation training for professional purposes (e.g. actors speech exercises, TV and radio anchors, PR personnel, etc.) and for treatment of various speech disorders, e.g. dyslexia.

To sum up, as far as language teaching is concerned, there are no technical limitations for further improvement of *Study Intonation* provided the language material corresponding to given learning level and conversational topic is properly selected.

6. References

- [1] D. M. Chun, "Computer-assisted pronunciation teaching," *The Encyclopedia of Applied Linguistics*, 2012.
- [2] T. M. Derwing and M. J. Munro, "Second language accent and pronunciation teaching: A research-based approach," *Tesol Quarterly*, vol. 39, no. 3, pp. 379–397, 2005.
- [3] J. B. Gilbert, *Teaching Pronunciation Using the Prosody Pyramid*. Cambridge: Cambridge University Press, 2008.
- [4] M. Teshima, "Nihon no chugakkoukotogakkou niokeru eigo no onseikyoiuiku nitsuite-hatsuonshido no genjo to kadai-(on teaching english pronunciation in japanese secondary education: How it is and how it should be)," *Journal of the Phonetic Society of Japan*, vol. 15, pp. 1–31, 2011.
- [5] R. Cauldwell, "Streaming speech: Listening and pronunciation for advanced learners of english," 2002.
- [6] J. Dalby and D. Kewley-Port, "Design features of three computer-based speech training systems," *The path of speech technologies in computer assisted language learning*, pp. 155–73, 2008.
- [7] K. Precoda and H. Bratt, "Perceptual underpinnings of automatic pronunciation assessment," *The path of speech technologies in computer assisted language learning*, pp. 71–84, 2008.
- [8] E. Estebas-Vilaplana, "The teaching and learning of 12 english intonation in a distance education environment: TL.tobi vs. the traditional models," *Linguistica*, vol. 57, no. 1, pp. 73–91, 2017.
- [9] F. TALLEVI, "Teaching english prosody and pronunciation to italian speakers: the kaspar approach," 2017.

- [10] D. M. Chun, "Computer-assisted language learning," *Handbook of research in second language teaching and learning*, vol. 2, pp. 663–680, 2011.
- [11] D. Chun, B. Smith, and R. Kern, "Technology in language use, language teaching, and language learning," *The Modern Language Journal*, vol. 100, no. S1, pp. 64–80, 2016.
- [12] D. M. Chun, "The role of technology in sla research," *Language Learning & Technology*, vol. 20, no. 2, pp. 98–115, 2016.
- [13] G. Molholt and F. Hwu, "Visualization of speech patterns for language learning," *The path of speech technologies in computer assisted language learning*, pp. 91–122, 2008.
- [14] C. Cucchiaroni and H. Strik, "Second language learners' spoken discourse: Practice and corrective feedback through automatic speech recognition," in *Smart Technologies: Breakthroughs in Research and Practice*. IGI Global, 2018, pp. 367–389.
- [15] J. Levis and L. Pickering, "Teaching intonation in discourse using speech visualization technology," *System*, vol. 32, no. 4, pp. 505–524, 2004.
- [16] M. Sánchez-Tornel and P. Pérez-Paredes, "Teaching and researching computer-assisted language learning," *ReCALL: the Journal of EUROCALL*, vol. 23, no. 2, p. 181, 2011.
- [17] Y. Lezhenin, A. Lamtev, V. Dyachkov, E. Boitsova, K. Vylegzhanina, and N. Bogach, "Study intonation: Mobile environment for prosody teaching," in *Cybernetics (CYBCONF), 2017 3rd IEEE International Conference on*. IEEE, 2017, pp. 1–2.
- [18] A. Monfared, S. E. Cervantes, S. M. Lee, and M. Jackson, "Establishing a praxis between sla theory and call-based practices," in *Applications of CALL Theory in ESL and EFL Environments*. IGI Global, 2018, pp. 59–79.
- [19] J. Six, O. Cornelis, and M. Leman, "Tarsosdsp, a real-time audio processing framework in java," in *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society, 2014.
- [20] E. Barkhudarova and F. Pankov, *Speak Russian with good pronunciation: a practical course of speech: textbool for foreign students (in Russian)*. Russky yazyk, 2008.