Cross-Platform Mobile CALL Environment for Pronunciation Teaching and Learning

Andrei Kuznetsov[1,*], *Anton* Lamtev[2,**], *Iurii* Lezhenin[2,***], *Artem* Zhuikov[2], *Mikhail* Maltsev[2] and *Elena* Boitsova[2] and *Natalia* Bogach[2], and *Evgeny* Pyshkin[3,****]

[1]JetBrains s.r.o.
[2]St.-Petersburg Polytechnic University, Russia
[3]University of Aizu

Abstract. Mobile technologies promote computer-assisted language learning (CALL) while mobile applications, being learner-oriented by design, provide a powerful founding to build individual self-paced environments for language study. Mobile CALL (MALL) tools are able to offer new educational contexts and fix, at least, partially, the problems of previous generations of CALL software. Nonetheless, mobile technologies alone are not able to respond to CALL challenges without cooperation and interaction with language theory and pedagogy. To facilitate and formalize this interaction several criteria sets for CALL software has been worked out in recent years. That is why an approach based on using mobile devices is a natural way to transfer the learning process from teaching-centered classroom to a process, which is oriented to individual learners and groups of learners with better emphasis on supporting individual learning styles, user collaboration and different teaching strategies. Pronunciation teaching technology in one of areas, where the automated speech processing algorithms and corresponding software meet the problems of practical phonology. Computer-assisted prosody teaching (CAPT), a sub-domain of CALL, is a relatively new topic of interest for computer scientists and software developers. Present-day advancement of mobile CAPT tools is supported by evolutionary processes in the theory of language learning and teaching. This paper explores language–technology relations using a case of StudyIntonation – a cross-platform multi-functional mobile CAPT tool based on a digital processing core for speech processing, visualization and estimation developed by the authors. We particularly address the problems of developing CAPT evaluation frameworks. To define the problematic points of the project and understand the directions for future work, we discuss an approach to formalized evaluation using a set of CAPT-specific criteria drawing attention to such evaluation factors as general descriptive information, instructional purposes, functionality, usability, and presentation.

1 Introduction

Mobile technologies for language study have gained much attention in recent years for their potential benefits for teachers and learners. Computer assisted language learning (CALL) tools developed for mobile devices (MALL) have strongly promoted their popularity due to the ability to foster a stress-free and self-paced learning environment, virtually unlimited input, and rich multimodal feedback [1].

In context of pronunciation training, an immediate and diverse feedback is especially beneficial. A wide scope of pronunciation teaching technologies and software aggregate speech processing algorithms and apply them in pronunciation pedagogy and assessment. Speech production patterns used to be detected by special instruments (e.g., CSL) or acoustic software (e.g., PRAAT) [2]. Speech science has been largely integrated into identifying various pronunciation features and favors either research or teaching (e.g., [3], [4]). The inventory of CALL tools to support learners through the use of a variety of multimodal feedback possibilities including sound, graphics, video, animation, spectral and pitch displays [5] becomes much more flexible and rich when implemented on a mobile platform.

Computer-assisted prosody teaching (CAPT), a subdomain of CALL, is a relatively new topic of interest for computer scientists and software developers. Automatic speech recognition (ASR) used pronunciation instruction is an important technology making a measurable impact on CALL by enabling the identification of particular parameters of the learner's output [6]. In turn, the advancement of mobile CAPT is supported by evolutionary processes in the theory of language teaching. Research on the role pronunciation plays for the overall language competency (e.g., [7], [8]) and points out the emergence of a number of shifts in instructional focus. The focus is moving from teacher-oriented to learner-oriented classrooms; the advancement of communicative and pragmatic goals of teaching over the abstract language learning [9] favors

*e-mail: andrei.kuznetsov@jetbrains.com
**e-mail: antonlamtev@gmail.com
***e-mail: lezhenin@kspt.icc.spbstu.ru
****e-mail: pyshe@u-aizu.ac.jp

the appearance of more concern about pronunciation. Pronunciation training is no longer considered a minor field; it is understood as one of the principle pivot points of successful communication [10].

By their design paradigm, mobile solutions support and encourage user personalizing and collaboration. That is why an approach based on using mobile devices is a natural way to transfer the learning process from teaching-centered classrooms to a process oriented to both individual learners and groups of learners with better emphasis on supporting individual learning styles and strategies. Moreover, mobile language learning tools can provide a private and flexible environment for self-study. The latter issue is particularly important for pronunciation learning, which is not mere acquisition of a new different phonological system, but mastering a complex cognitive skill of an individual [10].

Being in a rich, informal, contextual, and ubiquitous learning environment, in which it is possible for students to control their learning and receive diverse multimodal authentic input with tailored instructive feedback is highly motivating [11]. Students raise their awareness to improve pronunciation, become more conscious about their faculty of language and acquire a better understanding of many phonological features of a foreign language [1]. With mobile approach, learners gain a larger opportunity to act as active co-creators in cognition-driven process of pronunciation training.

In [12], the authors argue that the tool's availability is the primary factor of technology usability. Hubbard [13] names the foremost CALL principles to follow as access, authenticity, interaction and feedback. Unfortunately, even scientifically sound and engaging CAPT products often have a long way to everyday L2 pedagogy, while direct interaction with learners is the primary practical task of CAPT tools. Mobile CAPT tools can bridge this gap, making the instructional products more handy, available and intuitive to use. But so far, mobile CAPT software has not occupied its due place in teaching practice among other computer-based instructional products. The reasons of it might be attributed to three different sources: personal attitude of teachers and learners; developmental and algorithmic issues; lack of common evaluation guidelines.

In [14], Bitner et al. name eight areas of important consideration to form teachers' positive attitude to integrate technology into the curriculum. Even being inserted in teaching practice, CALL software might be used quite differently from what developers and teachers intended [15]. Human–computer interaction tracking research has shown that there is much individual learner variability in interaction with CALL programs [16], [17], [18], [19]. Kim [20] even reports that from the college EFL learner's perception, MALL are a sort of "subway-time study". Subsequently, the expectations of MALL content, style, and design are different from the concept of effective MALL in the literature and even far distant from effective CALL. Still an intrinsic learner motivation may origin from the possibilities that technology can offer. According to multiple evidence ([21], [22], [23]), the major drawback of existing CALL systems, and CAPT systems in particular, is a lack of guidelines how to employ technology to make courseware.

Although formal criteria, containing development agenda and benchmark sets, have been worked out to evaluate CALL, they are largely based on general software evaluation frameworks (e.g., [24], [25], [13]) and does not mainly address prosody but other language aspects such as computer assisted learning of vocabulary or grammar. While the frameworks for CALL evaluation are far from being inconsistent, none of them explicitly describe workflows for conducting evaluations or, moreover, articulate any guidelines for CAPT software development [26].

CAPT evaluation frameworks have been created only in very recent years. In [27] a checklist of criteria for CAPT software, websites, and mobile apps is proposed; a question-based instrument in [5] evaluates to what extent the software programs teach pronunciation in accordance with the principles of the communicative approach [28]. Having accumulated the previous background in CALL evaluation, these instruments concisely and explicitly inform CAPT developers about the expectations and priorities of CAPT target audience.

Based on the recent new release of StudyIntonation CAPT software [29], [30], [31], in this paper we discus the results of its screening [32] against the criteria from [27], in order to evaluate StudyIntonation functionality and disclose the limitations of its design.

2 Method

StudyIntonation software was conceived as a free open-source mobile application to provide language learners with the visuals of phrasal intonation [29]. Its availability for teachers and learners was discussed in [30]. StudyIntonation approach suggests the collaboration between teachers and native speakers in co-creation of the learning content and supposes various learning scenarios with respect to the learning style of a student.

StudyIntonation workflow (Figure 1) implements all the stages of prosody processing, beginning from pitch detection and approximation up to pitch visualization and pitch estimation. The system incorporates the modules for pitch detection and pitch approximation enabling processing both the model (native speaker) and the record (learner) pitch patterns, as well as for displaying the pitch curves in the client application. The module for pitch estimation is necessary for providing a primary feedback to learners. It can also be used for formalization of pitch similarity evaluation.

Our previous efforts led us to the implementation of a prototype for StudyIntonation developed for Android OS. The similar functionality was later implemented for Apple iOS as an independent application. The efforts towards more similarity of both versions resulted in the latest version which is a cross-platform CAPT environment that works identically under Android and Apple iOS. Both StudyIntonation implementations share the same digital signal processing core and can be used for both university education and self-training. As StudyIntonation has now evolved into a full-fledged mobile CAPT, cross-platform

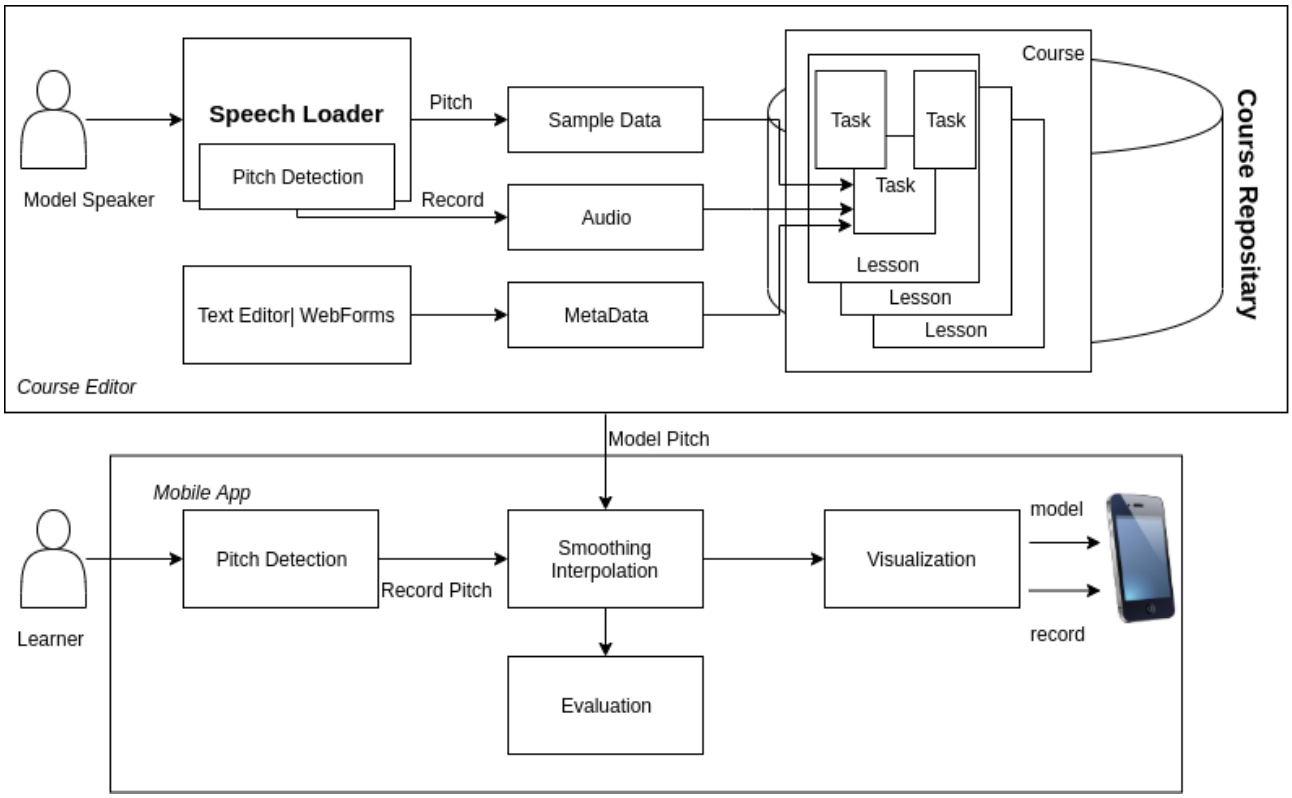

Figure 1. StudyIntonation Workflow

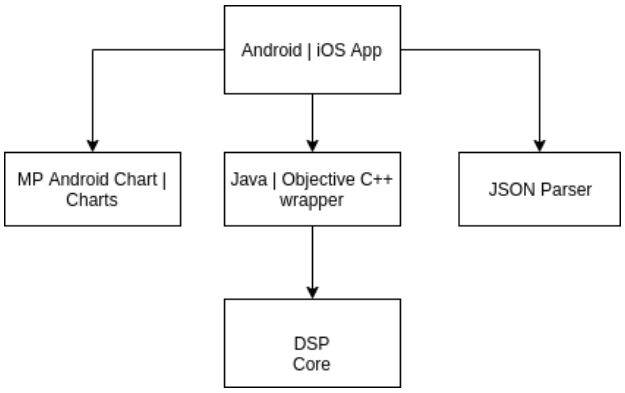

Figure 2. Cross-platform functionality support

and multi-functional, it needs to be validated with respect to the latest formal methods of CAPT tools evaluation, e.g., [27] to understand its soundness for linguistics and pedagogy.

3 Cross-Platform Implementation and DSP Core Assessment

Cross-platform implementation (Figure 2) does not change the workflow. Cross-platform functionality was mostly achieved through the choice of the development instruments and libraries existing for both target operating systems. For example, pitch graphics is processed and rendered by MP Android Chart and iOS Charts, the similar approach was used for the other functional components (such as json parsers, etc.).

In contrast to the the first prototype [29] where Java TarsosDSP was used for signal processing components, for the latest release, the DSPCore was rewritten in C++, thus, both applications share the same DSPCore library for pitch processing, and the pitches look identically (Figures 3, 4). This library is incorporated into Android and iOS applications through Java and Objective C++ wrappers respectively.

For fundamental frequency processing (signal interpolation, resampling, low-pass filtering), DSPCore uses a number of third-party components from Essentia open library of tools for audio data. The project leverages the ability to build Essentia library with a direct indication to the set of algorithms used. In this case, the algorithms not engaged into pitch processing are not included in the output file. Thus, the final output file of the DSPCore library has smaller size and weaker external library dependencies compared to the full assembly.

The DSPCore library API includes the following important functions. The function *bool loadAudioFile()* reads audio data from the specified file; *bool detectPitch()* calculates the fundamental frequency vector by using YIN algorithm; *bool smoothPitch()* is responsible for making the pitch signal more uniform; *float dtw()* implements the DTW time sequence comparison.

DSPCore has the following external dependencies: *fftw3f* for calculating the Fourier transform (FFT), and

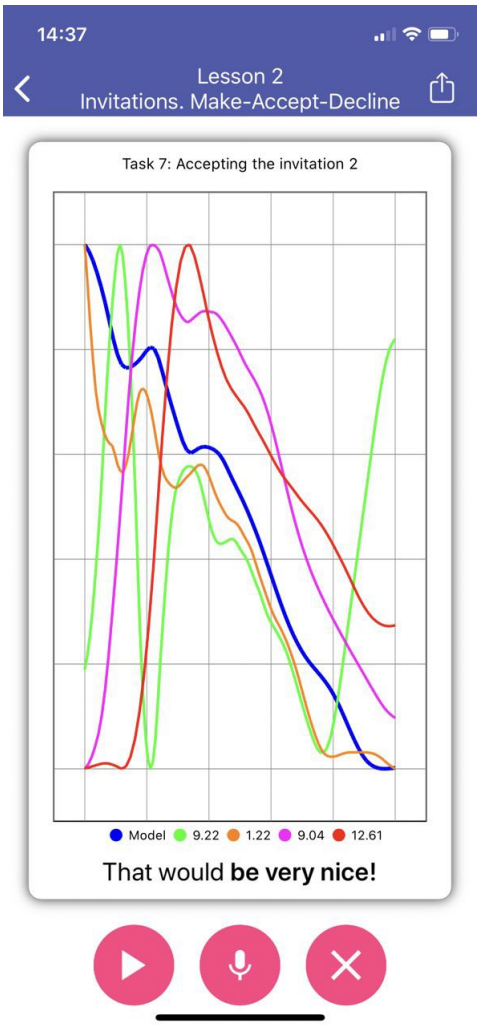

Figure 3. Application Appearance under Apple iOS (task interface)

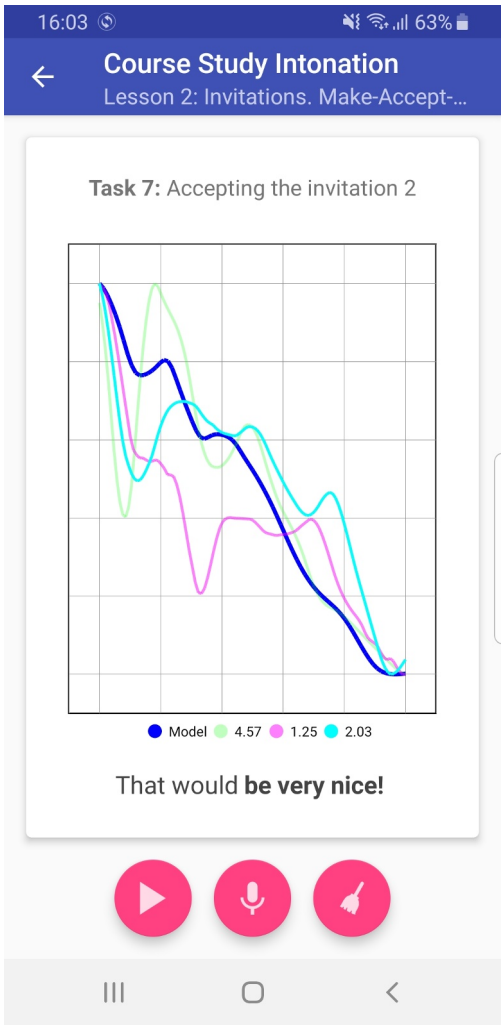

Figure 4. Application Appearance under Android OS task interface

samplerate for converting the sampling audio data frequencies.

To read audio data from a file, *ffmpeg* libraries are used; *avcodec* library implements audio codecs; *avformat* library implements streaming protocols, container formats, as well as basic input/output operations; *avutil* library implements some auxiliary functions; *avresample* library is used for audio mixing and resampling.

To test the DSPCore function, we used the prosodically labelled corpus IviE. Intonation contours in audio recordings of IViE speech stimuli were loaded into PRAAT and DSPcore. For each IViE record bearing the tones indicated in tone markup (H*, L+H*, L%, 0%, etc.), the corresponding tone movements were observed in both programs.

4 Criteria Evaluation and Discussion

Searching for a thorough professional analysis of the software product, we assess the proposed architecture as per criteria set introduced in [27], which are focusing attention on the characteristics that need to be considered when evaluating CAPT software. This set of criteria is particularly helpful because it draws attention to the factors that might be otherwise ignored. Therefore, it enables getting quick CAPT functionality summary and setting the goals for future development.

Henrichsen [27] organized the criteria in five sections: (A) general descriptive information, (B) instructional purpose(s) and activities, (C) functionality and usability, (D) instructional factors, and (E) presentation. The tables 1,2 contain the results of how StudyIntonation underwent the estimation with respect to this set of criteria.

C, D, and E contain aggregate quantitative measures, which are obtained through each of the items in sections C, D, and E rated according to the scale: -2 -1 0 1 2 (strongly disagree, disagree, no opinion, agree, strongly agree, NA=Does not apply, CT=Cannot tell (insufficient data)) (table 2).

In functionality and usability test (section C) StudyIntonation has got 14 points of 20 which were assigned for every positive answer. 6 points were subtracted since there is no operational “Help” for users (except the initial tutorial) nor a possibility to contact the creators (the terms

Section.Parameter	Value
A.Platform	iOS, Android
A.Language	British English
A.Instruction Language	English, Russian, German
A.Cost	Free
A.Target Audience	Any
A.Level of language	Any
B.Primary Objective	Intonation learning
B.Secondary Objective	Tone movement practice
B.Aspects of Pronunciation Addressed	Suprasegmentals
B. Activities	Listen, record, repeat
B.Feedback and Record Keeping	Instant, visual

Table 1. CAPT evaluation criteria. Parts A, B

Section	Score
C.Functionality and usability	14/20
D.Instructional factors	35/42
E.User interface and presentation	13/16

Table 2. CAPT evaluation criteria. Parts C, D, E

from [27] are used hereafter to name the features of the product under evaluation).

With respect to instructional factors (section D) StudyIntonation has got 35 out of 42 points. Positions where points were subtracted are related to our hesitation about the adequacy of pitch visualization and the feedback quality (-2 points in total). Multi-speaker mode, the choice of pronunciation features to work on and assignment diversity still requires further efforts (it resulted in -5 points). Just like for the acquisition of sounds, especially, of native language, multiple-talker models seem to be particularly efficient to improve perception of novel contrasts as the inherent variability allows for induction of general phonetic categories [33].

The score 13 (out of 16) was given in section E for user interface and presentation quality. 3 deduction points are due to the difficulty of user interface aesthetics and usability validation. 2 points were deduced because of the current inability to play audio at different speed.

5 Conclusion

Prosody models and tonal pattern classification are beneficial for both language education and research purposes, including speech processing an annotation algorithms based on the hypothesis that the informative parts of speech are usually prosodically highlighted by a speaker and have discernible characteristics [34]. That is why our approach addresses three target groups of users (namely, language learners, language teachers, as well as the researchers working in the domains of ASR and language phonology). The approach connects four basic ideas:

1. Harnessing multimodality in language learning (e.g., perception of interactive audio-visual channels) enforced by multimedia features of mobile devices;
2. Supporting different learning styles of students (e.g., visual, auditory and kinesthetic);
3. Advancing signal and speech processing, visualization and estimation algorithms to enhance technology-driven education and research;
4. Developing mobile tools leveraging rich existing experience of portable device users.

Our research continues the works (see, for example, the extensive studies [35],[36]) on developing approaches to extending audio perception by using visual perception channels represented as pitch graphs demonstrated to learners. Intelligent CALL using ASR algorithms becomes one of key innovations in language learning requiring more personalized and flexible tools for perceptual training [37]. However, it is important to mention that the CAPT tool usability factors are of equal importance as technical features (such as intelligent pitch processing techniques, ASR, pitch visualization, etc.).

Though there are important features, which are still missing in the current version of StudyIntonation (such as playback at different speed, speech context representation, more interactive feedback display), we may conclude that the evaluation frameworks (similar to one demonstrated in this paper), create an important environment for further collaboration between technical society and language teachers. Such a collaboration is one of ultimate priorities of present-day human centered technology and education development.

References

[1] M.C. Pennington, P. Rogerson-Revell, in *English Pronunciation Teaching and Research* (Springer, 2019), pp. 235–286

[2] O. Kang, *Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency*, in *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (2012), pp. 10–15

[3] H. Obari, H. Kojima, S. Itahashi, Glasgow, 10-13 July 2013 Papers p. 245 (2013)
[4] H. Obari, H. Kojima, *The Effect of Blended and Flipped Lessons on L2 Learning Using Mobile Technologies*, in *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (Association for the Advancement of Computing in Education (AACE), 2015), pp. 449–454
[5] C.G.d.F.M. Martins, J.M. Levis, V.A.M.C. Borges, Ilha do Desterro **69**, 141 (2016)
[6] E.M. Golonka, A.R. Bowles, V.M. Frank, D.L. Richardson, S. Freynik, Computer assisted language learning **27**, 70 (2014)
[7] T.M. Derwing, M.J. Munro, TESOL quarterly **39**, 379 (2005)
[8] J. Jenkins, Annual review of applied linguistics **24**, 109 (2004)
[9] J. Morley, TESOL quarterly **25**, 481 (1991)
[10] P. Trofimovich, *Interactive alignment: Implications for the teaching and learning of second language pronunciation*, in *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (2012), pp. 1–9
[11] H. Obari, S. Lambacher, *Successful EFL teaching using mobile technologies in a flipped classroom*, in *Critical CALL–Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (Research-publishing. net, 2015), pp. 433–438
[12] D. Chun, R. Kern, B. Smith, The Modern Language Journal **100**, 64 (2016)
[13] P. Hubbard, Language Teacher Education and Technology: Approaches and Practices p. 153 (2017)
[14] N. Bitner, J. Bitner, Journal of technology and teacher education **10**, 95 (2002)
[15] J.T. Pujolă, ReCALL **14**, 235 (2002)
[16] D.M. Chun, CALICO Journal **30**, 256 (2013)
[17] D.M. Chun, J.S. Payne, System **32**, 481 (2004)
[18] J. Collentine, Language Learning & Technology **3**, 44 (2000)
[19] T. Heift, M. Schulze, *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues* (Routledge, 2007)
[20] H. Kim, Glasgow, 10-13 July 2013 Papers **138** (2013)
[21] C. Chapelle, Language Learning & Technology **1**, 19 (1997)
[22] J.F. Jones, ELT Journal **55**, 360 (2001)
[23] M.C. Pennington, Computer Assisted Language Learning **12**, 427 (1999)
[24] C.A. Chapelle et al., *Computer applications in second language acquisition* (Cambridge University Press, 2001)
[25] J. Leakey, An integrated approach to effectiveness research in CALL. Bern: Peter Lang (2011)
[26] B.L. McMurry, D.D. Williams, P.J. Rich, K.J. Hartshorn, TESL-EJ **20**, n2 (2016)
[27] L. Henrichsen, *A System for Analyzing and Evaluating Computer-Assisted Second-Language Pronunciation-Teaching Websites and Mobile Apps*, in *Society for Information Technology & Teacher Education International Conference* (Association for the Advancement of Computing in Education (AACE), 2019), pp. 709–714
[28] M. Celce-Murcia, D.M. Brinton, J.M. Goodwin, *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide* (Cambridge University Press, 2010)
[29] Y. Lezhenin, A. Lamtev, V. Dyachkov, E. Boitsova, K. Vylegzhanina, N. Bogach, *Study intonation: Mobile environment for prosody teaching*, in *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)* (IEEE, 2017), pp. 1–2
[30] E. Boitsova, E. Pyshkin, Y. Takako, N. Bogach, I. Lezhenin, A. Lamtev, V. Diachkov, *StudyIntonation courseware kit for EFL prosody teaching*, in *Proc. 9th International Conference on Speech Prosody 2018* (2018), pp. 413–417
[31] N. Bogach, *Languages and cognition: towards new CALL*, in *Proceedings of the 3rd International Conference on Applications in Information Technology* (ACM, 2018), pp. 6–8
[32] T.M. Derwing, M.J. Munro, *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*, Vol. 42 (John Benjamins Publishing Company, 2015)
[33] A. Neri, C. Cucchiarini, H. Strik, L. Boves, Computer assisted language learning **15**, 441 (2002)
[34] A. Batliner, B. Möbius, in *The integration of phonetic knowledge in speech technology* (Springer, 2005), pp. 21–44
[35] E. Estebas-Vilaplana, Linguistica **57**, 73 (2017)
[36] C. Cucchiarini, H. Strik, in *Smart Technologies: Breakthroughs in Research and Practice* (IGI Global, 2018), pp. 367–389
[37] M. Qian, E. Chukharev-Hudilainen, J. Levis, Language Learning & Technology **22**, 69 (2018)