# ESTOÑOL, A COMPUTER-ASSISTED PRONUNCIATION TRAINING TOOL FOR SPANISH L1 SPEAKERS TO IMPROVE THE PRONUNCIATION AND PERCEPTION OF ESTONIAN VOWELS

**Katrin Leppik[1] and Cristian Tejedor-García[2]**
*[1]University of Tartu and [2]University of Valladolid*

**Abstract.** Over the past few years the number of online language teaching materials for non-native speakers of Estonian has increased. However, they focus mainly on vocabulary and pay little attention to pronunciation. In this study we introduce a computer-assisted pronunciation training tool, Estoñol, developed to help native speakers of Spanish to train their perception and production of Estonian vowels. The tool's training program involves seven vowel contrasts, /i-y/, /u-y/, /ɑ-o/, /ɑ-æ/, /e-æ/, /o-ø/, and /o-ɤ/, which have proven to be difficult for native speakers of Spanish. The training activities include theoretical videos and four training modes (exposure, discrimination, pronunciation, and mixed) in every lesson. The tool is integrated into a pre/post-test design experiment with native speakers of Spanish and Estonian to assess the language learners' perception and production improvement. It is expected that the tool will have a positive effect on the results, as has been shown in previous studies using similar methodology.

**Keywords:** CAPT tool, L2, production, perception, vowels, Estonian, Spanish, Estoñol

## 1. Introduction

In the past few years the number of foreigners living in Estonia has grown rapidly. According to Statistics Estonia (SA 2019), in January 2019 there were 532 native speakers of Spanish living in Estonia, which is 344 more speakers than in 2011 (REL 2011). The Ministry of Culture and the Ministry of Education and Research offer free Estonian courses for newcomers. In addition, there are some online courses and materials for (independent) studying of Estonian, for example, Keeleklikk (Kingisepp and Ilves), Keeletee (Kingisepp and Ilves), Kultuuriklikk (Kõivupuu and Rüütli), Speakly (Keskpaik and Ojamets), and Lingvist (Müntel and Jalakas 2013). In the Google Play store there are many

applications that offer Estonian, among other languages. Examples include Simply Learn Estonian (Simya Solutions), Speak Estonian: Learn Estonian Language Offline (Edutainment Ventures), and Learn Estonian (MSSVCORP).

These materials and tools are free and easily accessible, but they mainly focus on vocabulary, grammar, and listening, and pay very little attention to pronunciation. The pronunciation of Estonian is normally taught in word context, and very few instructions for the pronunciation are given.

On the other hand, some materials do focus on pronunciation; for example the handbook of Estonian pronunciation *Eesti keele hääl-damine* (Kraut 2000) includes detailed explanations with figures, exercises, and audio examples. An online material *Eesti keele ja kultuuri kursused* offers listen and repeat type of activities and sound categorization exercises (Rammo and Teral). Applications such as Learn Estonian – 50 languages (50 Languages) and Learn Estonian Free (MetaLanguage) pay some attention to pronunciation and provide the alphabet together with phonetic transcription and audio examples. Learn Estonian (Edu Master Pro) offers the learner the opportunity to record oneself and listen to the recording, but the application does not offer corrective feedback for the pronunciation.

All these courses, materials, and tools are designed mainly for English and Russian speakers and also pay some attention to German, French, and Finnish speakers. In second language (L2) acquisition, the learners' mother tongue (L1) plays a very important role. L2 acquisition models suggest that producing and perceiving sounds that are similar to the learners' L1 is more difficult than acquiring those that differ from L1 sounds. Two such models are the Speech Learning Model (Flege 1995) and the Perceptual Assimilation Model (Best 1995, Best and Tyler 2007). The Native Cardinality Method (NCM) is developed specifically for Spanish L1 speakers in order to improve the production of English vowels. NCM takes the students' native phonological system as a starting point and goes through an intensive cyclic training protocol that includes phases of articulatory knowledge (exposure), perceptive awareness (discrimination), and sound realization (production). NCM intends to avoid or reduce the transfer of L1 pronunciation to L2 (Cámara-Arenas 2010, 2014). NCM is included as the training

protocol applied to the computer-assisted pronunciation training system introduced in this paper.

Estonian and Spanish differ in the quantity and quality of vowels. Estonian has nine vowels /i, y, e, ø, æ, ɑ, o, ɤ, u/ while Spanish has five /i, e, a, o, u/ (Asu and Teras 2009 and Hualde 2005). The vowels /i, e, o, u/ are identical in both languages. The Estonian vowel /ɑ/ is an open back vowel, and it is produced more back than the Spanish open front vowel /a/. Additionally, Estonian has vowels /y, ø, æ, ɤ/ which do not have counterparts in the Spanish vowel system.

Studies focusing on the pronunciation and perception of Estonian vowels have shown that depending on the similarity of L1 and L2 there are some vowel sounds (/ø, ɤ, y, æ, ɑ/) that are difficult to acquire (Näätänen et al. 1997, Meister and Meister 2011, and Nemoto et al. 2015). Studies of Spanish L1 speakers' perception (Leppik 2017) and production (Leppik and Lippus 2014 and Leppik et al. 2019) show a similar tendency: (1) Spanish L1 speakers do not perceive the difference between sounds /ø/ and /ɤ/, (2) vowels /ø/ and /ɤ/ are merged together and produced as an ambiguous mid-vowel which overlaps to some extent with /o/, (3) Estonian /ɑ/ and /æ/ are produced as Spanish /a/, and (4) Estonian /y/ has some overlap with /u/.

Although traditional in-classroom courses are still a valuable and useful resource for L2 learning, one-to-one tutoring and e-learning are also attractive alternatives. In particular, Computer-Assisted Language Learning (CALL) and Mobile-Assisted Language Learning (MALL) systems integrate advanced technology that has become very interesting to language learning by helping the process of learning and teaching in an efficient way. Experts predict that there will be on average seven network connected smart devices (mobile phones, tablets, smart watches, computers, etc.) per person around the world in 2020 (Statista 2016). It is well known that these devices cannot substitute for human tutors, though they can perform a complementary role in education by increasing motivation and efficiency in the learning process, as they can be used anywhere, anytime, and can provide as many repetitions of the training unit as desired.

Computer-Aided (assisted) Pronunciation Training (CAPT) is one of the most important sub-areas of CALL and MALL, constantly under-going rapid change. It combines automatic pronunciation quality evaluation and corrective feedback, among other functionalities provided

by speech technology systems. Two main speech technologies are Automatic Speech Recognition (ASR) and Text-To-Speech (TTS), which transform speech into written text, and vice versa, respectively. Currently these systems are supported by complex algorithms and an enormous quantity of data that improve their quality significantly. For instance, recent advances in machine learning processing have helped Google to generate audio waveforms 1,000 times faster than before (50 milliseconds for generating a one-second audio) and have achieved quality ratings over 20% better than standard voices (Lardinois 2018). Furthermore, Google has also achieved a word accuracy rate of 95% for speech recognition of English, therefore reaching the threshold of human accuracy (Meeker 2017). Better rates are likely to be obtained in the near future, not only by Google, but also by other companies and institutions, with the use of more powerful central processing units and more sophisticated deep neural networks (Prabhavalkar et al. 2017 and Zhang et al. 2017).

Thomson and Derwing (2014) reported that as of 2014 only 20 studies about CAPT had been published in peer-reviewed venues, mainly focused on English as a foreign language. However, CAPT systems are being gradually incorporated into educational experiments due to the improvements described in the previous paragraph and the new possibilities they offer (Katz and Assmann 2019).

Previous works have reported on the design of a CAPT system for Spanish L1 speakers to practice the pronunciation of English as L2 (Tejedor-García et al. 2017) and the results of an experiment with Japanese L1 speakers learning Spanish as L2 (Tejedor-García et al. 2018a, 2018b). In both cases, the CAPT system and the experiment were based on the minimal-pairs technique and NCM, and the experimental design included exposure, discrimination, and production tasks. The potential of such a design for learner engagement and its possibilities for generating assessments of user performance were measured. This paper introduces a computer-assisted pronunciation training tool called Estoñol, which is designed for Spanish L1 speakers to help train their perception and production of Estonian vowels. Following and expanding upon previously mentioned experiments, novel perception tests and feedback resources are included, particularly adapted for Estonian as L2 and Spanish as L1.

## 2. Description of the CAPT tool

### 2.1. Material of the CAPT tool

Estoñol[1] is a vowel perception and production training system based on minimal-pairs tasks. A minimal pair is a pair of two words that differ in only one sound. The tool uses similar methodology as described by Tejedor-García et al. (2017, 2018a, and 2018b). Based on the results of previous perception (Leppik 2017) and production studies (Leppik and Lippus 2014 and Leppik et al. 2019) a list of seven vowel contrasts (/i-y, u-y, ɑ-o, ɑ-æ, e-æ, o-ø, o-ɤ/) was composed to train Spanish L1 speakers' perception and production of Estonian vowels. For every vowel contrast, 12 minimal pairs are presented to the language learners (e.g., *karu–käru*, "bear–cart"; *kare–käre*, "rough–hoarse"; *ranne–ränne,* "wrist–migration"*; kagu–kägu,* "southeast–cuckoo"*; sara–sära*, "barn–shine"). The minimal pairs were selected from the dictionary *Eesti õigekeelsussõnaraamat ÕS 2018* (ÕS 2018) using regular expressions; words including difficult-to-pronounce combinations for Spanish L1 speakers were excluded from the selection.

All the minimal pairs of the tool were tested following a protocol developed for similar tools like Japañol (Tejedor-García et al. 2018b) and TipTopTalk! (Tejedor-García et al. 2016) to ensure that the speech synthesizer and recognition utilities manage the material of the tool without problems. First, the pairs were tested with the synthesizer EKI kõnesüntesaator[2] (Mihkla et al. 2012 and Eesti Keele Instituut 2017), and only well-synthesized words (i.e., those with native-like pronunciation, good quality, correct stress, and quantity) were included in the tool. Subsequently, the well-synthesized pairs were tested with the speech recognition utility Kõnele[3] (Alumäe and Tilk 2018 and Kaljurand 2018). The pairs were produced five times in different manner (different intonation, speech rate, loudness) by a native speaker of Estonian, and only pairs recognized five or four times out of the five repetitions were included in the tool.

Due to technological limitations (speech synthesis and speech recognition), the words included in the tool have varying structures (e.g.,

---

1   Available at https://eca-simm.uva.es/es/proyectos/capt/estonol/
2   https://play.google.com/store/apps/details?id=ee.eki.ekisynt
3   https://play.google.com/store/apps/details?id=ee.ioc.phon.android.speak

*olu–õlu,* "circumstance–beer"*; noor–nöör,* "young–string"*; toru–tõru,* "pipe–acorn"*; tekk–täkk,* "blanket–stallion"), but they are all maximally trisyllabic (e.g., *looming–lööming,* "creation–brawl"*; tootlus–töötlus,* "production–processing"*; sorkima–sörkima,* "to poke–to jog"). Some of the words are not very common in Estonian (e.g., *perm–pärm,* "Perm–yeast"*; kimmel–kümmel*, "roan–cumin liqueur"), but this is not regarded as a problem because they were included with the aim of teaching the vocalic sound and not the meaning of the word. In the minimal pairs list there are nouns, adjectives, and verbs in different forms (e.g., *tappis–täppis,* "to kill 2SG PST–exact"*; loodi–löödi,* "to create PST PASSIVE–to hit PST PASSIVE"*; sule–süle,* "feather SG GEN–lap SG GEN").

Each lesson of the tool starts with a theoretical mode (all the other modes will be explained in the next section). This mode combines textual and audiovisual components which explain the pronunciation of the sounds and highlight the main differences. Pronunciation is also illustrated with examples (text and sound) and short videos explaining the articulatory movements of each vowel. The next training exercises are monitored by the system via interaction log files. They consist of entries that contain information about correct/incorrect production events, correct/incorrect perception events, a list of speech recognition hypotheses and their confidence scores, number of TTS uses by the learners, points achieved, and timing events.

## 2.2. Stages of the CAPT tool

Estoñol is a CAPT tool for Android-supported smart devices with an adaptive design. Figure 1 presents the main stages of the application. Once the application is installed on the device, users must log in to the system by typing their authorized email (stage 1). After logging in, learners select the lesson (unit) to practice (stage 2). Each lesson contains a different minimal-pair contrast. The score reached by the user (expressed as a percentage) is regularly updated on the screen. Lessons must be undertaken in consecutive order; therefore the link to lesson 2 becomes active only when lesson 1 is successfully completed, and so forth.
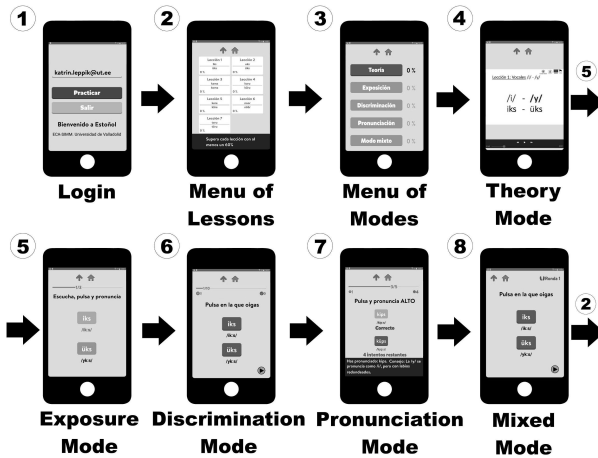
**Figure 1.** Screenshots of all training stages of the Estoñol CAPT tool.

In the third stage, the sequence of training modes based on the NCM is displayed after choosing the lesson. In order to be completed, each lesson takes the user through theoretical, exposure, discrimination, and production activities, a priori, in a strictly consecutive order. A final mixed mode is presented at the end of each lesson, where discrimination and production tasks alternate randomly. Exposure, discrimination, production, and mixed modes must be performed with a success rate of at least 60% (random responses alone would generate a rate of 50%). Otherwise, learners have the opportunity to retry the failed mode or to enter the mode recommended by the system, based on the user's results. For instance, if a user reaches a 50% success rate in the discrimination mode, the system will recommend retrying the exposure mode.

Each training mode contains a fixed number of tasks with minimal pairs (three in exposure, 10 in discrimination, 10 in production, and nine in the mixed mode). Strict control is ensured by the system because neither lessons, nor modes within lessons, nor tasks within training modes, can be skipped or undertaken in an order other than the one established by the CAPT tool. Also, each training mode is accessible by clicking on a corresponding button on the menu of modes, when they are enabled.

The theory mode is the first training mode (stage 4) where a short multimedia video presents concepts and tips on the articulation of

sounds of the minimal pair of the lesson, pointing out the similarities and differences between L1 and L2. The option to advance to the next mode becomes available only at the end of the video. Within the training session, learners may choose to review the material as many times as they want. At stage 5 (exposure mode), the preliminary exposure to the contrasts presented in the theoretical video is reinforced. In this mode, users must listen, repeat, and compare three minimal pairs, without limitations on the number of attempts. This mode serves as a feedback recommendation by the system, when users get bad results in the next training modes. In stage 6 (discrimination mode), users must identify the word produced by the TTS in each of the 10 tasks; the tasks are maximally 10 seconds long. Participants are allowed to listen to the synthesized model of the words as many times as they want. Its speech rate alternates between normal and slow. Stage 7 is the production training mode. There is a limit of five attempts per word and 60 seconds per pair. Additionally, after three consecutive failures, the system executes an explicit corrective feedback response that recommends to users that they listen to the synthesized version of the problematic word. Furthermore, the tool displays short feedback tips after incorrect production events; for example, "The vowel /y/ is pronounced like /i/ but with rounded lips". Mixed mode is the final training mode (stage 8), which works as a review activity, because it incorporates again both discrimination and pronunciation tasks. In this mode, four perception and five production tasks alternate randomly, summing up a total of nine task activities.

## 3. Discussion

### 3.1. Advantages of the tool

Estoñol was designed to help Spanish L1 speakers train their production and perception of Estonian vowels. We believe that language learning programs that are designed for a specific L1 group are more effective than those that do not take into consideration the learners' L1. Comparing the language learners' L1 and L2 helps to pinpoint the difficult aspects that need more attention and training. L2 acquisition models (Flege 1995, Best 1995, Best and Tyler 2007, and Kuhl 1991) support the idea that language learners' L1 has a significant role in the L2 acquisition process. The tool is expected to have positive perception

and production improvement results comparable to those achieved in similar studies for other languages (Tejedor-García et al. 2017, 2018a, 2018b, and 2019-submitted).

Previous studies have proven that CAPT tools have some benefits as compared to traditional language learning methods, for both learners and teachers. They include interaction-rich and stress-free contexts where teachers have more opportunities to attend to the individual needs of students, because not all situations can be predicted and programmed by a computer algorithm, while the students can practice at their own pace and get real-time personalized feedback (Levy and Stockwell 2013). In particular, ASR-based CAPT systems present several advantages, such as dynamic evaluation, more intensive practice, immediate-automated feedback, anxiety-free context, individualized feedback, and opportunities for repair (Neri et al. 2002, 2006). Even though old TTS systems generated a great deal of controversy when used as a pedagogical tool, recent research in speech synthesis has reported some benefits in terms of naturalness, comprehensibility, accuracy, and intelligibility (Smith et al. 2015 and Bione et al. 2016).

The Estoñol tool uses TTS and ASR, which have many advantages. TTS allows us to use words without pre-recording them, and it makes it very easy to renew the tool's material whenever needed – which may simply be an updated list of words. Also, the learners can record their pronunciation and listen and compare it with TTS. ASR benefits the learners when giving instant feedback on their pronunciation, which is very important in order for the learner to achieve a more native-like pronunciation of L2.

For the purposes of research, this tool can be used to gather material and record speakers independent of their location. The use of mobile and smart devices gives us the opportunity to include more participants in our studies, as the participants record themselves with smart devices and do not need to come to the university to participate in a study. The majority of participants to date have been exchange students who came to Estonia for a term or academic year and who have studied Estonian for a term or two. The tool helps us find participants with different backgrounds (age, language experience, country of origin), which is important to get a better overview of the Spanish L1 speakers' pronunciation of Estonian.

### 3.2. Further developments and limitations

In addition to the differences in vowel inventory, Estonian and Spanish also differ in consonantal inventory, and there are some combinations that are difficult for beginners to pronounce. Estonian is also rich in diphthongs that are challenging for L2 speakers. The tool could be improved by adding lists containing consonant sound contrasts and diphthongs. To add to the lists, the procedure explained in section 2.1. should be carried out.

The Estonian language is one of the few languages that has a three-way quantity system consisting of short, long, and overlong quantity degrees which function over a disyllabic foot. Quantity in Estonian is realised by combining the durational and tonal components (see, e.g., Lehiste 1960). The production and perception of Estonian long and overlong quantity degrees is difficult for non-native speakers (for details see Meister et al. 2015). The Estonian quantity contrasts could be added to the tool as a further development. To add the quantity contrasts, a new approach is needed because the speech synthesizer (Mihkla et al. 2012 and Eesti Keele Instituut 2017) has some trouble producing isolated words in long and overlong quantity. As a solution, pre-recorded words could be used, which would necessitate some changes in the software. Alternatively, the Estonian quantity could be taught in sentence context, as the speech synthesizer is better at producing the quantity in context than in isolated words.

A similar experiment as that conducted by Tejedor-García et al. (2018b) is planned in order to evaluate the effect of the tool on participants' perception and production of Estonian vowels. The experiment consists of training sessions and pre- and post-testing, for which the results will be analysed. In previous studies (Tejedor-García et al. 2017, 2018a, 2018b, and 2019), training sessions with the tool have shown a positive effect, and in the case of Estonian vowels such improvement is expected as well.

## 4. Conclusions

There are many language learning courses and applications to help non-native speakers learn Estonian, but they mainly focus on vocabulary and grammar. Unfortunately, there are very few Estonian learning

applications that explicitly teach pronunciation, and the ones that do so do not offer corrective feedback. To fill the gap in the field, the CAPT tool Estoñol was created. Estoñol has a theoretical training mode that explains the articulatory movement of the vowels, followed by different types of exercises (exposure, discrimination, pronunciation, and mixed mode). Previous studies (Tejedor-García et al. 2017, 2018a, 2018b, and 2019-submitted) using similar technology and methodology have shown a positive effect of the tool on the production and perception of L2 sounds, and this is also expected in the case of Estonian. Estoñol could be improved by adding Estonian quantity, consonant sound contrasts, and diphthongs to the tool.

CAPT systems provide innovative practices and research that lead to transforming language learning, developing opportunities to revisit old ideas, and challenging established beliefs. Estoñol has many benefits for both language learners and researchers. The tool uses ASR, which gives the language learner an opportunity to receive corrective feedback, and CAPT systems provide an anxiety-free environment for language learning. Thanks to TTS, it is easy to update the tool's material, and the use of mobile and smart devices helps researchers to find more participants for their studies.

**Addresses:**
Katrin Leppik
Institute of Estonian and General Linguistics
University of Tartu
Jakobi 2–409
51005 Tartu, Estonia
E-mail: katrin.leppik@ut.ee

Cristian Tejedor-García
Department of Computer Science
University of Valladolid
Paseo de Belén 15, Despacho 2L018
47012 Valladolid, Spain
E-mail: cristian@infor.uva.es

## References

50 Languages (s.a.) *Learn Estonian – 50 languages and learn Estonian*. Version 11.2. Available online at <https://play.google.com/store/apps/details?id=com.goethe.et>. Accessed on 27.05.2019.

Alumäe, Tanel, Ottokar Tilk, and Asadullah (2018) "Advanced rich transcription system for Estonian speech". In Kadri Muischnek and Külli Müürisep, eds. *Human language technologies – the Baltic perspective*, 1–8. (Frontiers in Artificial Intelligence and Applications, 307.) Amsterdam: IOS Press.

Asu, Eva Liina and Pire Teras (2009) "Illustrations of the IPA: Estonian". *Journal of the International Phonetic Association* 39, 3, 367–372.

Best, Catherine T. (1995) "A direct realist view of cross-language speech perception". In W. Strange, ed. *Speech perception and linguistic experience: theoretical and methodological issues in cross-language speech research*, 167–200. Timonium: York Press.

Best, Catherine T. and Michael D. Tyler (2007) "Nonnative and second-language speech perception: commonalities and complementarities". In Murray J. Munro and Ocke-Schwen Bohn, eds. *Language experience in second language speech learning: in honor of James Emil Flege*, 13–34. Amsterdam: John Benjamins Publishing.

Bione, Tiago, Jennica Grimshaw, and Walcir Cardoso (2016) "An evaluation of text-to-speech synthesizers in the foreign language classroom: learners' perceptions". In Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny, eds. *CALL communities and culture–short papers from EUROCALL 2016,* 50–54.

Cámara-Arenas, Enrique (2010) "Native cardinality: on teaching American English vowels to Spanish students". In L. Pérez, I. Parrado, and P. Tabarés, eds. *Estudios de la Metodología de la Lengua Inglesa* 5, 155–174. Valladolid: Centro Buendía.

Cámara-Arenas, Enrique (2014) "The NCM and the reprogramming of latent phonological systems: a bilingual approach to the teaching of English sounds to Spanish students". *Procedia – Social and Behavioral Sciences* 116, 3044–3048. https://doi.org/10.1016/j.sbspro.2014.01.704

Edu Master Pro (s.a.) *Learn Estonian*. Version 1.1.1. Available online at <https://play.google.com/store/apps/details?id=com.asiedu.lang.learnestonian>. Accessed on 27.05.2019.

Edutainment Ventures (s.a.) *Speak Estonian: learn Estonian language offline*. Version 1.0.4. Available online at <https://play.google.com/store/apps/details?id=com.eduven.ld.lang.estonian>. Accessed on 27.05.2019.

Eesti Keele Instituut (2017) *EKI kõnesüntesaator.* Version 1.1.03. Available online at <https://play.google.com/store/apps/details?id=ee.eki.ekisynt>. Accessed on 27.05.2019.

Flege, James Emil (1995) "Second-language speech learning: theory, findings, and problems". In Winifred Strange, ed. *Speech perception and linguistic experience: Issues in cross-language research,* 233–277. Timonium: York Press.

Hualde, José Ignacio (2005) *The sounds of Spanish with audio CD*. Cambridge: Cambridge University Press.

Kaljurand, Kaarel (2018) *Kõnele*. Version 1.6.98. Available online at <https://play.google.com/store/apps/details?id=ee.ioc.phon.android.speak&hl=et>. Accessed on 27.05.2019.

Katz, William F. and Peter F. Assmann (2019) *The Routledge handbook of phonetics*. New York: Routledge.

Keskpaik, Ingel and Ott Ojamets (s.a.) *Speakly*. Available online at <www.speakly.me>. Accessed on 27.05.2019.

Kingisepp, Leelo and Marju Ilves (s.a.) *Keeleklikk*. Available online at <https://www.keeleklikk.ee>. Accessed on 27.05.2019.

Kingisepp, Leelo and Marju Ilves (s.a.) *Keeletee*. Available online at <https://www.keeletee.ee>. Accessed on 27.05.2019.

Kraut, Einar (2000) *Eesti keele hääldamine: käsiraamat harjutuste ja helinäidetega.* Ellen Niit, ed. Tallinn: TEA Kirjastus.

Kõivupuu, Marju and Malle Rüütli (s.a.) *Kultuuriklikk*. Available online at <https://www.keeleklikk.ee/en/learningsessions/KUL01_en/open>. Accessed on 27.05.2019.

Kuhl, Patricia K. (1991) "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not". *Perception & Psychophysics* 50, 2, 93–107. https://doi.org/10.3758/BF03212211

Lardinois, Frederic (2018) *Google Cloud launches a new text-to-speech engine for developers*. Available online at <https://techcrunch.com/2018/03/27/google-cloud-launches-its-new-text-to-speech-engine-for-developers/>. Accessed on 20.06.2019.

Lehiste, Ilse (1960) "Segmental and syllabic quantity in Estonian". In *American studies in Uralic linguistics*, 21–82. Bloomington: Indiana University Publications.

Leppik, Katrin (2017) *Eesti keele vokaalikategooriad hispaania ja eesti emakeelega katseisikute tajus*. Magistritöö. Tartu: Tartu Ülikool. Eesti ja üldkeeleteaduse instituut.

Leppik, Katrin and Pärtel Lippus (2014) "A comparative study of Estonian and Spanish vowels in L1 and L2 production". In *XXVIII Fonetiikan päivät. Turku 25-26. lokakuuta 2013. Konferenssijulkaisu,* 19–26. *Turku: Turun yliopisto.*

Leppik, Katrin, Pärtel Lippus, and Eva Liina Asu (2019) "The production of Estonian vowels in three quantity degrees by Spanish L1 speakers". In Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, eds. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 1154–1158. Canberra, Australia: Australasian Speech Science and Technology Association.

Levy, Mike and Glenn Stockwell (2013) "CALL dimensions: options and issues in computer-assisted language learning". *The Modern Language Journal* 91, 4, 723–725. https://doi.org/10.1111/j.1540-4781.2007.00639_25.x

Meeker, Mary (2017) *Internet trends 2017 – code conference.* Available online at <https://www.kleinerperkins.com/perspectives/internet-trends-report-2017>. Accessed on 20.06.2019.

Meister, Einar, Rena Nemoto, and Lya Meister (2015) "Production of Estonian quantity contrasts by Japanese speakers". *Journal of Estonian and Finna-Ugric Linguistics.* (special issue *Aspects of speech studies*) 6, 3, 79–96. https://doi.org/10.12697/jeful.2015.6.3.03

Meister, Lya and Einar Meister (2011) "Production and perception of Estonian vowels by native and non-native speakers". *Proceedings of Interspeech 2011,* 1145–1148.

MetaLanguage (s.a) *Learn Estonian Free.* Version 3.1. Available online at <https://play.google.com/store/apps/details?id=com.metalanguage.learnestonianfree>. Accessed on 27.05.2019.

Mihkla, Meelis, Indrek Hein, Mari-Liis Kalvik, Indrek Kiissel, Risto Sirts, and Kairi Tamuri (2012) "Estonian speech synthesis: applications and challenges / Sintez reči èstonskogo jazyka: primenenie I vyzovy". In Andrej A. Kibrik, ed. *Computational linguistics and intellectual technologies. Papers from the Annual International Conference "Dialogue" (2012)*, 443–453. Moskva: RGGU.

MSSVCORP (s.a.) *Learn Estonian.* Version 2.4. Available online at <https://play.google.com/store/apps/details?id=com.codepro.learnestonian>. Accessed on 27.05.2019.

Müntel, Mait and Ott Jalakas (2013) *Lingvist.* Available online at <https://lingvist.com/>. Accessed on 27.05.2019.

Näätänen, Risto, Anne Lehtokoski, Mietta Lennes, Marie Cheour, Minna Huotilainen, Antti Iivonen, Martti Vainio, Paavo Alku, Risto J. Ilmoniemi, Aavo Luuk, Jüri Allik, Janne Sinkkonen, and Kimmo Alho (1997) "Language-specific phoneme representations revealed by electric and magnetic brain responses". *Nature* 385, 6615, 432–434. https://doi.org/10.1038/385432a0

Nemoto, Rena, Einar Meister, and Lya Meister (2015) "Production of Estonian vowels by Japanese subjects". *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, [1–5].

Neri, Ambra, Catia Cucchiarini, and Helmer Strik (2006) "Selecting segmental errors in non-native Dutch for optimal pronunciation training". *IRAL-International Review of Applied Linguistics in Language Teaching* 44, 4, 357–404.

Neri, Ambra, Catia Cucchiarini, Helmer Strik, and Lou Boves (2002) "The pedagogy–technology interface in computer assisted pronunciation training". *Computer Assisted Language Learning* 15, 5, 441–467.

Prabhavalkar, Rohit, Kanishka Rao, Tara N. Sainath, Bo Li, Leif Johnson, and Navdeep Jaitly (2017) "A comparison of sequence-to-sequence models for speech recognition". *Proceedings of Interspeech 2017*, 939–943.

Rammo, Sirje, Maarika Teral (s.a) *Eesti keele ja kultuuri kursused. Hääldusharjutused algtasemel.* Tartu Ülikooli eesti keele (võõrkeelena) osakond. Available online at < https://keeleweb2.ut.ee/kursused/haeaeldusharjutused-algtasemele>. Accessed on 25.09.2019.

REL 2011 = Rahva ja eluruumide loendus. RLE06: loendatud püsielanikud emakeele ja soo järgi. Statistikaamet". Available online at <https://www.stat.ee/rel2011>. Accessed on 24.05.2019.

SA 2019 = Statistikaamet. Available online at <https://www.stat.ee/ee>. Accessed on 24.05.2019.

Simya Solutions Ltd. (s.a.) *Simply learn Estonian*. Version 4.1.10. Available online at <https://play.google.com/store/apps/details?id=simply.learn.estonian>. Accessed on 27.05.2019.

Smith, George, Walcir Cardoso, and César García Fuentes (2015) "Text-to-speech synthesizers: are they ready for the second language classroom?". *Proceedings of the Meeting on English Language Teaching (MELT) 2015.*

Statista (2016) *Number of network connected devices per person around the world from 2003 to 2020.* Available online at <https://www.statista.com/statistics/678739/forecast-on-connected-devices-per-person/>. Accessed on 20.06.2019.

Tejedor-García, Cristian, Valentín Cardeñoso-Payo, and David Escudero-Mancebo (2018a) "Japañol: a mobile application to help improving Spanish pronunciation by Japanese native speakers". *Proceedings of IberSPEECH 2018*, 157–158.

Tejedor-García, Cristian, Valentín Cardeñoso-Payo, María J. Machuca, David Escudero-Mancebo, Antonio Ríos, and Takuya Kimura (2018b) "Improving pronunciation of Spanish as a foreign language for L1 Japanese speakers with Japañol CAPT tool". *Proceedings of IberSPEECH 2018*, 97–101. https://doi.org/10.21437/IberSPEECH.2018-21

Tejedor-García, Cristian, David Escudero-Mancebo, Enrique Cámara-Arenas, César González-Ferreras, and Valentín Cardeñoso-Payo (2019, submitted) "Assessing pronunciation improvement in students of English using a controlled computer assisted pronunciation tool". Manuscript submitted for publication.

Tejedor-García, Cristian, David Escudero-Mancebo, César González-Ferreras, Enrique Cámara-Arenas, and Valentín Cardeñoso-Payo (2016) "TipTopTalk! Mobile application for speech training using minimal pairs and gamification". *Proceedings of IberSPEECH 2016,* 425–432.

Tejedor-García, Cristian, David Escudero-Mancebo, César González-Ferreras, Enrique Cámara-Arenas, and Valentín Cardeñoso-Payo (2017) "Evaluating the efficiency of synthetic voice for providing corrective feedback in a pronunciation training tool based on minimal pairs". *Proceedings of the 7th ISCA Workshop on Speech and Language Technology in Education,* 25–29.

Thomson, Ron I. and Tracey M. Derwing (2014) "The effectiveness of L2 pronunciation instruction: a narrative review". *Applied Linguistics* 36, 3, 326. https://doi.org/10.1093/applin/amu076

ÕS 2018 = *Eesti õigekeelsussõnaraamat ÕS 2018*. Maire Raadik, ed., Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik, comp. Tallinn, Eesti Keele Instituut: Eesti Keele Sihtasutus.

Zhang, Yu, William Chan, and Navdeep Jaitly (2017) "Very deep convolutional networks for end-to-end speech recognition". *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4845–4849.

**Kokkuvõte. Katrin Leppik ja Cristian Tejedor-García: Estoñol, mobiilirakendus hispaania emakeelega eesti keele õppijatele vokaalide häälduse ja taju treenimiseks.** Eesti keele õppimiseks on loodud mitmeid e-kursusi ja mobiilirakendusi, kuid need keskenduvad peamiselt sõnavara ja grammatika õpetamisele ning pööravad väga vähe tähelepanu hääldusele. Eesti keele häälduse omandamise lihtsustamiseks töötati välja mobiilirakendus Estoñol, mis on mõeldud hispaania emakeelega eesti keele õppijatele. Varasemad uurimused on näidanud, et hispaania emakeelega eesti keele õppijatele valmistab raskusi vokaalide /a, y, ø, æ, ɤ/ hääldamine. Mobiilirakenduse sisu on jagatud seitsmeks peatükiks, kus on võimalik harjutada vokaalipaaride /i-y/, /u-y/, /a-o/, /a-æ/, /e-æ/, /o-ø/, /o-ɤ/ tajumist ja hääldamist. Iga peatükk algab teoreetilise videoga, millele järgnevad taju- ja hääldusharjutused. Mobiilirakenduse mõju hindamiseks keeleõppija hääldusele ja tajule plaanitakse läbi viia eksperiment.

**Märksõnad:** CAPT, eesti keel, hispaania keel, L2, hääldus, taju, vokaalid, Estoñol