SpeakGreek: AN ONLINE SPEECH TRAINING TOOL FOR L2 PEDAGOGY AND CLINICAL INTERVENTION

Katerina Nicolaidis, George Papanikolaou, Evia Kainada, Konstantinos Avdelidis

Aristotle University of Thessaloniki
knicol@enl.auth.gr; pap@eng.auth.gr; ekainada@gmail.com; kon.avdel@gmail.com

ABSTRACT

This paper presents the first online biofeedback speech training tool for Greek designed to offer training on the production and perception of segmental and suprasegmental aspects of Greek. It is addressed to learners of Greek as an L2 and to clinical populations with articulation and phonation problems. The tool comprises four components: (i) the Phonetic Library, i.e., the acoustic, articulatory and visual description of different pronunciation aspects, (ii) the Basic Training, which trains users on phonation, pitch and intensity manipulation, (iii) the Sound Trainer, and (iv) the Melody Trainer, with perception and production exercises on sounds, stress and intonation. All components of production include real-time biofeedback. Initial results from L2 learners of Greek using two of the tool's components show improvement on production of the intonation of wh-questions and on consonant identification.

Keywords: speech training, pedagogy, clinical intervention, biofeedback, Greek.

1. INTRODUCTION

The key role that proficiency in oral language plays in communicative competence is undisputable. Despite common problems in oral production by L2 learners, pronunciation teaching has been largely marginalised in the language classroom, although recently it has steadily been gaining ground [6]. Articulation is also commonly targeted in speech therapy settings, given that a large percentage of communication difficulties can stem from articulation/phonological impairments [12].

Technological advances in the field of speech analysis with the use of computers and specialized software has led to the utilization of speech technologies for extra or alternative training in the L2 classroom and in speech therapy. The main rationale behind speech training devices is that through "speech visualization", that is, through the graphic representation of speech on a computer screen [11 (p.4), 23], elements of production become visible to the speaker and thus more easily distinguishable. Visualisation of feedback can take the form of *biofeedback*, i.e., the real-time moment-to-moment information about a physiological event, which can aid the user in gaining control over the component they are training on (e.g. real-time spectra/spectrograms and pitch contour displays, real-time displays of vowel production on an F1xF2 vowel space). Biofeedback has been applied in the remediation of numerous speech disorders, e.g., hearing impairment, articulation, fluency and voice disorders, e.g. [4, 20], and in L2 pronunciation teaching, e.g. [7]. Advanced computer-based biofeedback training tools often include an Automatic Speech Recognition (ASR) component that judges speaker's productions according to a stored database. ASR-derived feedback has been shown to improve pronunciation; however, concerns have been voiced for, e.g., low tolerance to acceptable non-native pronunciation or ineffectiveness in spontaneous speech environments [10, 11].

The tool presented in this paper incorporates linguistic knowledge to computer-based ASR and biofeedback components. It is the first online biofeedback speech training tool for Greek designed to respond to a growing demand for computer-based training in Greek (see also [24] which to our knowledge is not available). The need for a freely available, user-friendly, adaptive speech training system stems from two important changes in the language training scenery over the past decades: (a) Greek language teachers and therapists are steadily faced with larger student numbers and caseloads but significant time and resource constraints in their professional contexts, (b) learners of Greek as an L2 and individuals with speech disorders request the use of computer-assisted technology so as to benefit from additional individualised practice and training time, be autonomous and adapt to their different rates of learning, thereby improving overall performance, and reducing academic and professional failure, and social marginalization.

2. DESIGN PRINCIPLES

This section presents the theoretical decisions and research conducted for the development of the tool.

2.1. Theoretical framework

Key features included in the basic design of the training tool are:

(a) inclusion of exercises on perception and production. Training on perception has been shown to be beneficial for production [3, 14, 16, 17] and vice versa [5], highlighting the importance of tapping both sides of the speech mechanism.

(b) inclusion of exercises on segmental and suprasegmental aspects of Greek. Both segmental and suprasegmental training can affect comprehensibility, intelligibility and accentedness to different degrees, suggesting that both aspects need to be addressed, e.g. [1, 8, 9, 25].

(c) use of multi-talker samples for perception and production training. Research on high-variability phonetic training shows that input from multiple native speakers is beneficial [13,19, 25, 26].

(d) inclusion of exercises that are structured hierarchically; learners progress from isolated sounds, to syllables, to words and sentences. In line with the communicative approach for pronunciation teaching, this structure promotes incremental gains in perception and production by typically progressing from smaller to larger units and from controlled to freer activities [6].

(e) inclusion of exercises that target common pronunciation difficulties for Greek [18, 21, 22].

(g) different graphic interfaces and/or menus to cater for (i) different age groups, (ii) user profiles (learner vs. instructor), and (iii) the linguistically informed vs. naïve user.

2.2. Speech databases

Three speech databases were recorded and analyzed for the purposes of the tool. All were recorded using a Beyerdynamic MC 836 short shotgun cardioid lobe microphone writing directly on a desktop computer using a Nanoface sound card set at 44.100Hz sampling rate. The audiovisual samples were recorded with an HDC-Z10000 Digital Camcorder. Prompts were played back at the speakers using *ProRec* [15]. Segmentation was performed automatically and was then manually checked in PRAAT.

Speech database A provides data to the Phonetic Library (Section 3.1). Audio and video recordings were made from 3 Greek speakers (adult male and female, and child) producing all Greek vowels and consonants in isolation, in syllables and in words with the target sounds in different prosodic positions (e.g., /s/, /sa/, /'supa/, /'telos/, /'ðasos/). In addition, words with different stress patterns (ultimate, penultimate, antepenultimate, enclitic) and intonational contours (statements, polar questions, wh-questions, polite vs. impolite speech, continuation rises, etc.) were recorded. A total of 550 items were recorded per speaker. Electropalatographic data from one speaker and ultrasound data from two speakers were also recorded for the lingual sounds (312 items).

Speech database B is used for the acoustic analyses of sounds and for feeding the ASR component. Part of the material is also used for perception and production exercises. It includes audio data from 20 adult male, 20 adult female, and 20 children (10 boys and 10 girls, 8-10 years of age). It contains multiple repetitions of (a) isolated vowels, (b) sustained vowels and consonants, (c) vowels and consonants in syllabic frames (e.g., /pa/, /si/, etc.), (d) vowels and consonants in disyllabic words with stress on V1 or V2 (e.g., /'pata/, /pa'ta/), (d) selected longer words with/without enclitic stress (e.g., /'ðaskalos/, /'ðaska'los mu/), (e) short utterances with different focus position and intonational patterns (statements, wh-/yes-no questions, polite, impolite production), (f) a short sample of spontaneous speech. A total of 110.140 items were recorded for the segmental analyses and 12.420 for the suprasegmental analyses.

Speech database C provides the material for the perception and production exercises of the tool. Four speakers were recorded producing all words and sentences needed for the exercises on vowels, consonants, stress, and intonation (32.180 items).

2.3. Computational design

To achieve system accessibility and scalability, the implementation of the tool was organized in three parts (a) the speech database engine, (b) the analysis, biofeedback algorithms, and techniques, (c) the presentation engine.

The speech database engine is adapted to the operational needs of each component. For database A, it provides a web-based interface for the content management of each phonetic category (description, images, audio, video). Databases B and C contain audio plus metadata and are structured in an easy to maintain/extend file-based layout, available locally or remotely by web-services for query and access.

The techniques and algorithms include the analysis and biofeedback sections. The analysis procedures are designated to pull data from database B and either push back metadata information or extract expert knowledge to be incorporated into biofeedback procedures along with data from database C. Analysis is performed by local tools and/or suites (such as PRAAT or Matlab) while biofeedback is primarily Action Script 3.0 plus web

services based in order to comply with the needs of the presentation layer.

The presentation layer contains (a) passive HTML content which refers to the dynamic Phonetic Library elements framed by the static elements of general theming plus descriptive information, and (b) the training tools active content which was designed and implemented on the Adobe Flash infrastructure. The platform hosting all of the above is Drupal 7 CMS.

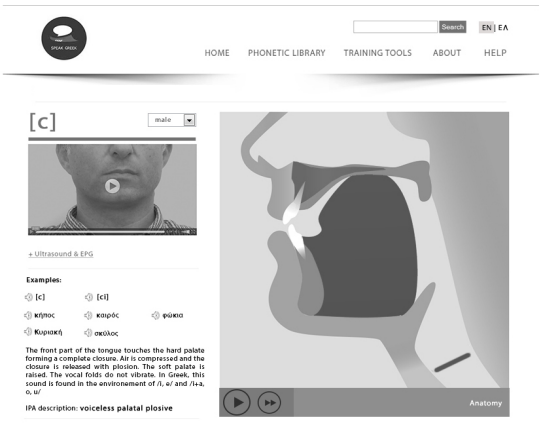
3. COMPONENTS OF THE TOOL

3.1. Phonetic library

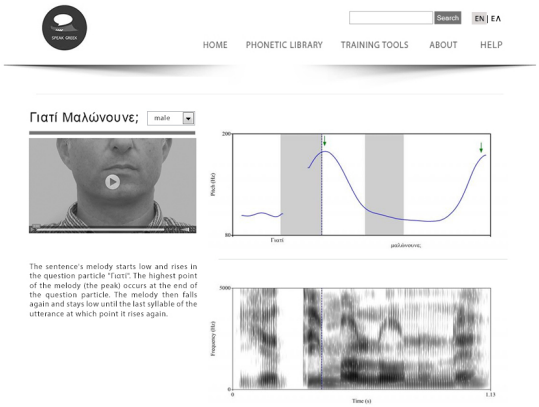
The library provides a description of (a) all Greek vowels and consonants presented with animated vocal tract diagrams (Fig. 1a), phonetic and articulatory description, audio and video recordings, electropalatographic and ultrasound data for selected items, (b) stress patterns (audio and video recordings), and (c) intonational patterns of utterances of varying type, length and focus presented with audio and video recordings, annotated F0 display and spectrogram (Fig. 1b).

Figure 1 a, b: Screenshot of the segmental (a) and the suprasegmental (b) components of the Phonetic Library.

(a)



(b)



3.2. Basic training

This component trains the user on: (i) phonation duration, (ii) frequency and amplitude control and modulation, (iii) frequency and amplitude range, and (iv) voicing on and off. Real time feedback on production is given in simple displays. For phonation duration, reference ranges (max vs. min phonation duration) were determined by the analysis of sustained vowels and consonants recorded in *Speech Database B*. For the frequency and amplitude applications, the user's habitual F0 and loudness levels are measured at the beginning of the exercise, and are then used to dynamically alter ranges (Fig. 2). Voicing on/off provides feedback on whether the speaker's productions are voiced or not.

Figure 2: Screenshot of a F0 modulation exercise. Users are required to start from their habitual F0 level and go over a pre-specified higher F0 level.

3.3. The sounds of my language

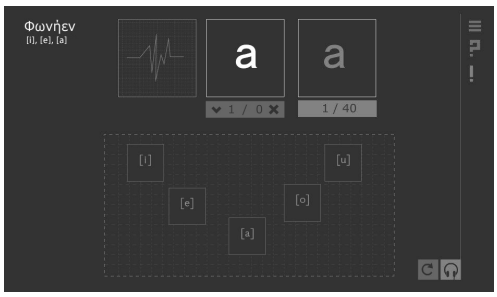
This component provides training on the perception and production of Greek sounds. Perception training is performed with the 'Sound Trainer'; the user initially selects the consonants or vowels s/he would like to practice. S/he then chooses specific segmental and prosodic contexts (e.g. position in word, minimal pairs). The specified listening material is chosen randomly from *Speech Databases B* and *C* and includes productions from multiple speakers. For minimal pairs, the system allows the user to select particular sound pairs and includes minimal pairs in isolation or in sentences (paradigmatic and syntagmatic). The user performs an identification task and immediately receives feedback (Fig. 3a, b).

Training on sound production is broken down into vowels and consonants, and users move from isolated sounds to sounds in syllables, words and sentences. Real-time feedback on vowels produced in isolation is given by plotting user productions on an F1xF2 vowel space (Fig. 4 a, b). Feedback on consonants produced in isolation is based on selected acoustic parameters and varies depending on consonant category (different feedback for e.g.,

fricatives vs. nasals). Normal ranges for measured parameters (e.g. vowel formants) are determined from the acoustic analyses of *Speech Database B*. They are dynamically altered by the system once the user specifies their gender and age. Feedback on segments produced in longer contexts relies on ASR. Feedback techniques for words and utterances are currently being developed.

Figure 3 a, b: Screenshot of the 'Sound Trainer' for vowels (a) and for consonants (b).

(a)



(b)

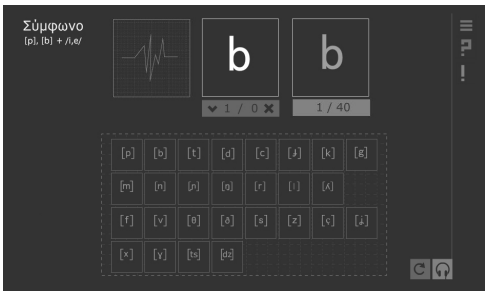
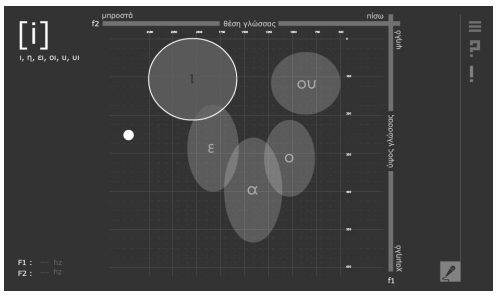
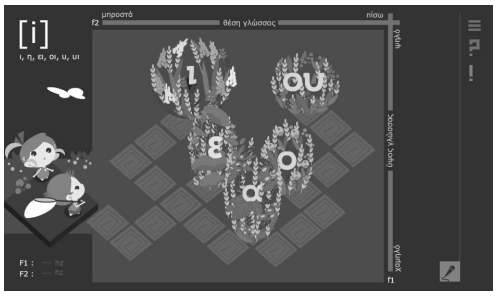


Figure 4 a, b: Screenshot of vowel space display for real-time feedback on isolated vowel production. Different graphics are illustrated in (a) and (b) for different age groups and individual preferences.

(a)



(b)



3.4. The melody of my language

At the suprasegmental level, the user receives training on production and perception with the 'Melody Trainer'. Perception training includes identification exercises on, e.g., stress position or intonational patterns. Similarly to the sounds of my language, the specified listening material is chosen randomly from the databases and includes productions from multiple speakers. For production training, the user can see a real-time display of their F0 overlaid over a target intonation contour which includes typical variation measured from the utterances of *Speech Database B*.

4. INITIAL EVALUATION

While the tool is under development, initial evaluation has been conducted. Six learners of Greek as a foreign language from mixed backgrounds were asked to use the Phonetic Library (audio and video recordings and F0 pitch displays) to train themselves on the intonation of wh-questions following a pre-defined protocol for five days. They listened to the same 35 wh-questions 6 times each day and paid attention to the F0 display provided. They were recorded producing two repetitions of 16 wh-questions before and after training. An intonational analysis of pitch targets using GrToBI [2] was performed, and utterances were classified as being correctly produced or not, that is as carrying the correct full set of intonational targets appearing at the correct segmental landmarks or not. This yielded a percentage of how many times each learner produced correct intonation before and after testing. Learners' productions improved by 32%. In addition, the 'Sound Trainer' was used by one English learner of Greek to train on the identification of /t, d/ (produced with short-lag vs. voicing lead in Greek) for four days consecutively, twice a day. She was also asked to use the Phonetic Library for an audiovisual illustration of the sounds. A 25% improvement in sound identification was achieved. Even though testing of the system is still at its very initial stages, these results, together with the positive feedback that was received from the learners, are encouraging for the effectiveness of the tool.

ACKNOWLEDGEMENTS

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)-Research Funding Program: ARISTEIA II. Investing in knowledge society through the

European Social Fund. Project title: "SpeakGreek: Developing a biofeedback speech training tool for Greek segmental and suprasegmental features: Application in L2 learning/teaching and clinical intervention" 3542. Thanks are due to all project team members and all the subjects recorded for the databases. We would also like to thank The Athens Center, P. Andreou and all the learners who used the speech training tool for its evaluation.

5. REFERENCES

[1] Anderson-Hsieh, J., Johnson, R., Koehler, K. 1992. The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning* 42, 529–555.

[2] Arvaniti, A., Baltazani, M. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In: Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press, 84-117.

[3] Bradlow A.R., Pisoni, D.B., Akahane-Yamada, R. Tokhura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am*. 101, 2299-2310.

[4] Brooks, S., Fallside, F., Gulian, E., Hinds, P. 1981. Teaching vowel articulation with the Computer Vowel Trainer: Methodology and results. *British Journal of Audiology,* 15, 151-163.

[5] Catford, J.C., Pisoni, D.B. 1970. Articulatory training in exotic sounds. *The Modern Language Journal*, 54(7), 477-481.

[6] Celce-Murcia, M., Brinton D.M., Goodwin, J. M. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.

[7] Chun, D.M. 2007. Technological advances in researching and teaching phonology. In: Pennington, M.C. (ed.), *Phonology in context.* Basingstoke: Palgrave Macmillan, 135-158.

[8] Derwing, T.M., Munro, M.J. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 39(3), 379-397.

[9] Derwing, T.M., Munro, M.J. 2009. Putting accent in its place: rethinking obstacles to communication. *Language Teaching* 42(4), 476-490.

[10] Franco, H., Bratt, H., Rossier, R., Gadde V. R., Shriberg, E., Abrash, V., Precoda, K. 2010. EduSpeak(R): A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing* 27, 401-418.

[11] Godwin-Jones, R. 2009. Emerging technologies-Speech tools and technologies. *Language Learning & Technology,* 13(3), 4-11.

[12] Harasty, J., Reed, V.A. 1994. The prevalence of speech and language impairment in two Sydney metropolitan schools. *Australian Journal of Human Communication Disorders,* 22, 1-23.

[13] Hardison, D.M. 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology* 8, 34–52.

[14] Hazan, V., Sennema, A., Iba, M., Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47, 360–378.

[15] Huckvale, M. 2012. ProRec: Speech Prompt and Record System [computer software, version 1.4]. http://www.phon.ucl.ac.uk/resource /prorec/

[16] Lambacher, S.G., Martens, W.L., Kakehi, K., Marasinghe, C.A., Molholt, G. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied PsychoLinguistics* 26, 227–247.

[17] Lengeris, A. 2009. Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica* 66, 169–187.

[18] Levanti, E., Kirpotin, L. Kardamisti, E., Kampouroglou, M. 1998. *I Fonologiki Ekseliksi ton Paidion stin Ellada* [Phonological acquisition of Greek children]. Athens: Greek Logopedic Association.

[19] Logan, J.S., S.E. Lively, Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. Acoust. Soc. Am*. 89, 874–886.

[20] Maryn, Y., de Bodt, M., van Cauwenberge, P. 2006. Effects of biofeedback in phonatory disorders and phonatory performance: A systematic literature review. *Applied Psychophysiology and Biofeedback,* 31(1), 65-83.

[21] Mennen I., Okalidou, A. 2006. Acquisition of Greek phonology: an overview. *Working Paper* 11, Queen Margaret University. http://eresearch. qmu.ac.uk/153

[22] Nicolaidis, K., Andreou, P., Bozonelos, V., Mavroudi, A., Theodorou, D., Tasioudi, M., Tsiantoula, S. 2011. Cross-linguistic influences in the acquisition of the phonetic/phonological system of Greek as a second/foreign language. *Proc 31 Annual Meeting of the Department of Linguistics*, Faculty of Philosophy, Aristotle University, Thessaloniki, 357-378.

[23] Olson, D.J. 2014. Benefits of visual feedback on segmental production in the L2 classroom. *Language and Learning Technology*, 18(3), 173-192.

[24] Öster, A-M., House, D., Protopapas, A., Hatzis, A. 2002. Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Paedia). *Proceedings of Fonetik, TMH-QPSR*, 44 (1), 45-48.

[25] Trofimovich, P., 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, 1–30.

[26] Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruiit, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., Nemoto, I. 2009. Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage* 46, 226-240.