AUDITORY CATEGORIES AND LARYNGOSCOPIC/ULTRASOUND IMAGES IN THE *iPA Phonetics* APP

John H. Esling [*1], Christopher Coey [†1] and Scott R. Moisik [‡2]
[1]Department of Linguistics, University of Victoria, Victoria, BC V8W 2Y2 Canada
[2]Max Planck Institute for Psycholinguistics, Nijmegen, 6525 XD, Netherlands

1 The sounds of the IPA in an App

iPA Phonetics is a self-contained iOS application created at the University of Victoria with the intention of illustrating the auditory categories and acoustic sound distinctions of the International Phonetic Alphabet (IPA) as well as the articulations that correspond to the production of those sounds. The App illustrates IPA consonant and vowel sound production with high-quality colour videos and illustrates an inventory of phonetic voice qualities with audio files whose labels are mapped onto a graphic that orients the categories geographically on the oral and laryngeal parts of the vocal tract. The format of the consonant chart in the App is an expanded elaboration of the official IPA chart, including many more characters and sound combinations, all of which are implied in phonetic theory. The vowel chart is also in IPA format, with one added symbol. The App gives users of Apple iOS mobile electronic devices (iPhone, iPad) the ability to hear, view, manipulate, and compare the sounds and articulations symbolized in the charts and voice quality table. For phonetic training, the App also has a 'matching game' function, which presents users with random video files to test their knowledge of the phonetic symbols and sounds. The videos are endoscopically filmed views of the oral vocal tract and of the laryngeal vocal tract. Laryngeally articulated sounds are also accompanied by a choice of three ultrasound views of the laryngeal vocal tract.

2 Purposes and functionality of the App

The purpose of this free App [1] is to introduce the general public as well as specialized users of IPA symbolization, via iPad/iPhone technology, to the auditory/acoustic inventory of possible speech sounds of the languages of the world and to how each sound is physically articulated. The App is a learning tool created in a university environment for the benefit of students, teachers, researchers, and the wider community at large. The format of the App is illustrated in Figure 1. By touching a symbol, each sound category can be listened to and viewed in the form of close-up orally-filmed endoscopic videos of the vocal tract. Images/audio may be sped up or slowed to view articulatory detail and gain an understanding of the phonetic classification system and the degrees of auditory distance between categories in the taxonomy. Videos of consonant and vowel categories can be expanded to full screen, with a slider bar for manual control. Both the consonant chart and vowel chart can toggle on/off the IPA Numbers of all symbols [4]; a useful resource for coders. The consonant chart has a zoom in/out feature.

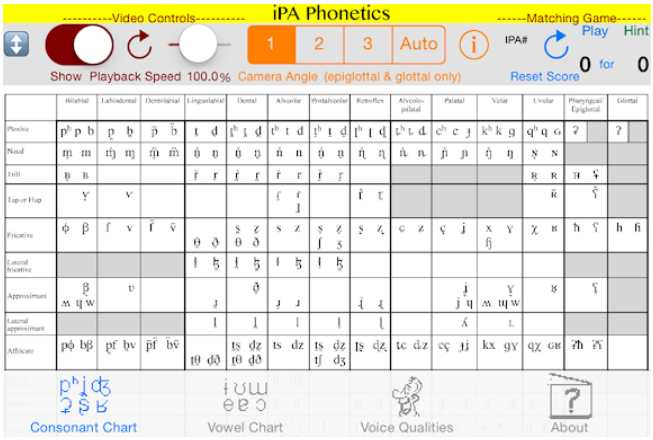

Figure 1: The main page of the *iPA Phonetics* App, showing a portion of the Consonant Chart, option selection bars, and links to the Vowel Chart and Voice Qualities pages and information page.

3 Auditory/Acoustic/Visual properties

3.1 Visual design

The chart format follows the elaborated inventory of speech sounds of the IPA published in the revised *Handbook of Phonetic Sciences* [2]. Categories are added to the standard set of IPA places and manners of articulation drawing on the ExtIPA chart [3,4] for articulations considered 'normal phonetic' possibilities as opposed to disordered/clinical, i.e. the Dentolabial and Linguolabial columns. An Alveolo-palatal column is added, expanded from its brief mention in the official IPA chart. Pharyngeal is considered identical in articulatory place to Epiglottal. There are four major divisions in the consonant chart (separated by bold lines): the first three columns are Labial; the second five, Front Lingual; the third four, Back Lingual; and the last two, Laryngeal. Symbols are included for as many cells in the matrix as phonetically practicable, given articulatory possibilities. Shaded areas denote phonetic impossibilities. Diacritic placement follows IPA principles. The Plosive row explicitly includes aspirated stops. The Nasal row is fully elaborated for all places with a voiceless counterpart in each cell. Trills involve the lips, tongue tip, uvula, and the aryepiglottic folds of the Pharyngeal articulator (using the IPA symbols for Epiglottal fricatives). Labiodental/Bilabial flaps, a Uvular tap (controversial in past IPA charts), and a newly designated Pharyngeal tap appear. Fricatives contrast rows for sibilance or grooving. Approximants include a set of 'approximated fricatives' [5]. Affricates are elaborated, in parallel to Fricatives, including for Pharyngeals. Implosives, Ejectives (with affricates), and Clicks (the sparse set from

the IPA chart rather than a full set [6]) are added as manners in this chart. The vowel chart, in the format of the 2005 IPA chart with one unrounded near-close back symbol added, has four playback options: Long, Short, Pairs, and 'sweeps' along various axes of the chart. This is a novel programming feature for phonetics illustrations, taking advantage of iOS capabilities.

The visual presentation of articulations in this App is distinguished by its use of endoscopic photography. Views of Labial sounds show the outside of the mouth, views of Lingual sounds look into the mouth, and views of Laryngeal sounds are taken with the rigid endoscope over the back of the tongue to see into the throat (larynx). Symbol selection, performance of the phonetic categories, and laryngoscopic filming were carried out by the first author. The design and programming of the App were accomplished by the second author. Simultaneous ultrasound data capture and graphic depiction of ultrasound/larynx views were done by the third author. Ultrasound images of the lower region of the vocal tract, previously never included in a database of articulatory categories, complement the lingual ultrasound images of oral sounds in other databases [7].

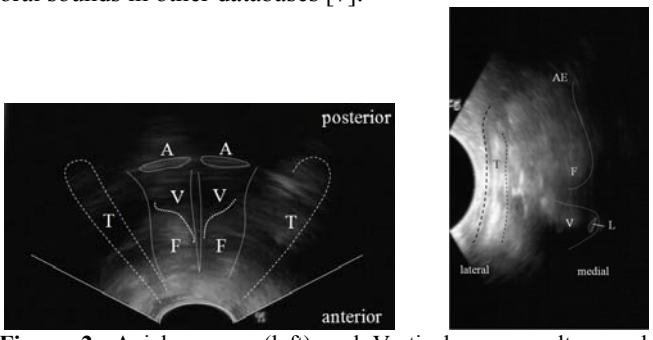

Figure 2: Axial convex (left) and Vertical convex ultrasound (right). A, arytenoid cartilages; V, vocal folds; T, thyroid cartilage; F, ventricular folds; L, vocal ligament; AE, aryepiglottic folds.

The App also has a Voice Qualities page with audio of oral and laryngeal categories, oriented on a static graphic of the vocal tract representing the Laryngeal Articulator Model of speech production [8]. Each category is listed by label and grouped into strategic articulatory regions, especially in the laryngeal vocal tract, as an exploratory feature for users.

3.2 Technical aspects

Video/audio clips for each set of symbols were captured using a KayPENTAX 9100 Rhino-Laryngeal-Stroboscope, 9105 70º-angle rigid oral laryngoscope (hand-held), a 35mm lens to view oral articulations and a 28mm wide-angle lens to view laryngeal articulations, connected via a Panasonic GP-US522 camera to multichannel-synchronizing software and procedures developed by the second author.

Ultrasound images were captured simultaneously with the laryngoscopic images of Laryngeals using a GE portable LOGIQe R5.0.1 system and 8C-RS convex probe to image supraglottal laryngeal involvement (e.g. for Glottal stop) and 12L-RS straight-line probe at a relatively shallow 2-4 cm depth on the neck and about 2-4 cm of the vertical dimension for clear resolution of laryngeal structures to image larynx height changes (e.g. for Pharyngeal/ Epiglottals). This is a novel laryngeal technique that differs from the approach usually taken in oral lingual ultrasound data capture. The three viewing options are an axial convex view (Figure 2), a vertical convex view (Figure 2), and a vertical flat (straight-line) view showing neck muscle movement. Anatomical sections of larynx models are included with corresponding labelled ultrasound images in an 'INFO' page, viewable via the toolbar, to illustrate the detail of what can be seen using laryngeal ultrasound [9].

4 Challenges in producing the resource

There are a number of technical and theoretical challenges in producing a resource of this nature. The first is to select phonetic categories that reflect as exhaustively as possible the auditory/acoustic separation between 'cardinal' speech sound values. The second is to perform these canonical values with accurate articulatory gestures and sound quality, while abating the noise of the laryngoscopy and ultrasound instruments in the audio signal, recalibrating the intensity of the light source and camera resolution, holding the scope (and ultrasound probe) in an optimal position for filming, then post-processing the video to obtain phonetically ideal auditory/acoustic distinctions and uniform video/colour and audio quality across all final files. The virtue of the iOS environment, from the point of view of technical implementation, is that novel approaches to displaying the phonetic subject matter can be seen and tested by a wide crowd of users, who then have a role to play in developing the effectiveness of the educational resource.

References

[1] Christopher Coey, John H. Esling, and Scott R. Moisik. 2014. *iPA Phonetics*, Version 1.0 [2014]. Department of Linguistics, University of Victoria.

[2] John H. Esling. 2010. Phonetic notation. In William J. Hardcastle, John Laver & Fiona E. Gibbon (Eds.), *The Handbook of Phonetic Sciences,* 2nd ed. Oxford: Wiley-Blackwell, 678-702.

[3] Martin J. Ball. 1991. Computer coding of the IPA: Extensions to the IPA. *JIPA,* 21:36-41.

[4] IPA. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

[5] Eugenio Martínez-Celdrán. 2004. Problems in the classification of approximants. *JIPA*, 34:201-210.

[6] Amanda L. Miller, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, and Chris Collins. 2009. Differences in airstream and posterior place of articulation among N|uu clicks. *JIPA,* 39:129-161.

[7] http://www.seeingspeech.arts.gla.ac.uk/uti/index1.html Carnegie Trust Ultrasound Tongue Imaging Resource.

[8] John H. Esling. 2005. There are no back vowels: The laryngeal articulator model. *Can. J. of Linguistics,* 50:13-44.

[9] Scott R. Moisik. 2013. *The Epilarynx in Speech*. PhD dissertation, University of Victoria.