Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Education in Acoustics
Session 3aED: Learning by Listening: Education in Acoustics Based on Listening (Lecture/Poster Session)

3aED2. Learning acoustic phonetics by listening, seeing, and touching

Takayuki Arai*

***Corresponding author's address: Dept. of Information and Communication Sciences, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, 102-8554, Tokyo, Japan, arai@sophia.ac.jp**

There is a huge volume of written textbooks available in virtually every modern field, including acoustic phonetics. However, in areas dealing with acoustics, learners often face problems and limitations when they deal with only written material and no audio or visual information. As one response to this problem, we have developed several sets of physical models of the human vocal tract and have shown that they are extremely useful for intuitive understanding. In addition, we also developed a tool called "Digital Pattern Playback." Another solution is an online version featuring demonstrations. We are currently collecting materials, mainly in the form of sounds, for educational purposes in acoustics and phonetics and are releasing them as "Acoustic-Phonetics Demonstrations" through our Web site. These demonstrations are designed for students in linguistics, phonetics and phonology, speech pathology, audiology, psychoacoustics, speech engineering, and others. However, potential users are not limited to these groups, as we feel that a wide range of learners can obtain tremendous benefits from the demonstrations, including those who are studying foreign languages or patients undergoing speech articulation therapy. [This work was partially supported by a Grant-in-Aid for Scientific Research (24501063) from the Japan Society for the Promotion of Science.]

Published by the Acoustical Society of America through the American Institute of Physics

©2013 Acoustical Society of America [DOI: 10.1121/1.4799629]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Different types of audio/visual tools have been developed for use in many fields including acoustics and phonetics. In addition to descriptions in textbooks, educational tools yielding inputs into our five senses help facilitate learners' intuitive understanding and are extremely effective in education. Today, as a consequence of computers and networks coming into widespread use, tools that make use of computers, multimedia, and the worldwide web (WWW) have become very popular throughout the world.

Because acoustics is, of course, a field involving sounds, many tools that make use of sound demonstrations are available for use in it. However, many acoustic phenomena are often invisible because vibration is too fast, because air particles as a medium cannot be seen, and for other reasons. Therefore, a number of tools have been developed that focus on the visualization of sounds. Thus, educational tools in acoustics and phonetics have a wide content range, including the visualization of invisible acoustic phenomena and the provision of speech sounds that are difficult to produce. These tools make use of actual images, computer simulation, computer graphics, animation, and many other things.

One of the noteworthy tools developed for use in the acoustics field is “Auditory Demonstrations”, which is provided in the form of a compact disk (CD) issued by the Acoustical Society of America (ASA) [1]. This CD deals with a variety of different topics, mainly in the auditory area, with explanatory descriptions covering such subtopics as critical band, sound pressure vs. loudness, masking, pitch, timber, beat/distortion/echo, and binaural effect.

We have been collecting a set of demonstrations similar to the Auditory Demonstrations, mainly in the area of acoustic phonetics. Our set is an online version featuring acoustic demonstrations [2, 3]. The materials are mainly in the form of sounds and videos with explanatory descriptions in acoustics and phonetics provided for educational purposes. We have released them as “Acoustic-Phonetics Demonstrations (APD)” through our website [4]. These demonstrations are designed for students in linguistics, phonetics and phonology, speech science speech pathology, audiology, psychoacoustics, speech engineering, and other fields. However, potential users are not limited to these groups, as we feel that a wide range of learners can obtain tremendous benefits from the demonstrations, including those who are studying foreign languages or patients undergoing speech articulation therapy.

One set of tools we have been developing intensively that is now a part of the APD is a group of various versions of vocal tract and related models (e.g., [5-8]). As an example of its use, learners often encounter difficulties understanding the acoustic theory of vowel production; for them, physical models of the human vocal tract are helpful in understanding the theory. These models are useful not only for students being trained in engineering but also for students in humanities, such as linguistics and psychology. By combining a lung model and a vocal-tract model, we can help learners to understand the whole speech production system from breathing to phonation and articulation. We have been applying them for children as well, and confirmed that the set of vocal-tract models is intuitive and helpful for teaching speech science to elementary, junior high, and high school students. An example of its application was seen at the 2012 science workshop at the National Museum of Nature and Science in Tokyo, Japan [9].

Another tool that we have developed, which is also a part of the APD, is the “Digital Pattern Playback (DPP)” [9-12] based on the “Pattern Playback (PP) developed by Cooper and his colleagues from Haskins Laboratories in the late 1940s for converting a spectrogram into sound [13-15]. The PP tool has contributed greatly to the development of research in speech science. We developed the digital version by using digital signal processing, so that the system could be implemented on a personal computer (PC). The DPP system consists of a video (or web) camera, a PC, and a loudspeaker with an amplifier. With this system, learners can listen to sounds reproduced from either a spectrogram from their own speech or from a pre-printed spectrogram.

In this paper, we describe some of the demonstrations in the APD and discuss further potential topics for a future APD.

ACOUSTIC-PHONETICS DEMONSTRATIONS

Table 1 shows a part of the list in the APD, which mainly covers acoustic phonetics and related areas such as articulatory phonetics, auditory phonetics, and psychoacoustics. It also covers basics in acoustics, digital signal processing, psychology, etc. In the following sections, the selected items will be explained in detail.

TABLE 1. A part of the list in the Acoustic-Phonetics Demonstrations.

Categories	Topics
Basic Acoustics	Propagation of sound
	Simple harmonic motion
	Period vs. aperiodic waves
	Standing waves
	Resonance
Digital Signal Processing	Sampling and quantization
	Discrete-time signal
	Linear time-invariant system
	Impulse response
	Filters
	Fourier analysis/synthesis
	Spectrum (FFT, LPC, cepstrum)
	Sound spectrogram
	Digital pattern playback (DPP)
Anatomy and Physiology	Lungs
	Larynx and phonation (modal/breathy/creaky)
	Pharynx (including nasopharyngeal port)
	Tongue
	Oral cavity / Nasal cavity
	Lips
Introduction to Acoustic Phonetics	Fundamental frequency
	Formants
	Source-filter theory
	Uniform tube model
	Two-tube model
	Three-tube model
	Perturbation theory
Vowels	Vowels in F1-F2 chart
	Vocal-tract models
	Simple vowels
	Diphthongs
	Nasalization
Stops	Acoustics
	F2-onset frequency and place of articulation
	Voice onset time (VOT)
Fricatives	Acoustics
	Sibilants (/sa/ vs. /sha/)
	Non-sibilants
Affricates	Acoustics
	Affricates vs. stops
	Affricates vs. fricatives (/shi/ vs. /chi/)

TABLE 2 (cont'd).

Categories	Topics
Nasals	Acoustics
	Place of articulation
	Hypernasality
Approximants	Acoustics
	Formant transitions
	English /r/ vs. /l/
	Japanese /r/ sounds
Other Consonants	Trills/taps/flaps
	Clicks
Acoustic Effects	Co-articulation
	Prosody
	Speaker variations
	Speech disorders
Basic Audition	Anatomy and physiology
	Frequency response and audiogram
	Middle-ear simulation
	Cochlea simulation
	Auditory deficits and hearing impairment
Auditory Psychophysics	Sensation of pitch
	Critical band and masking
	Temporal masking
	Binaural effects
Speech Perception	Introduction
	Formants (JND, locus theory, etc.)
	Categorical perception
	Top-down vs. bottom-up processing
	Cocktail party effect
	Delayed auditory feedback
	Phonemic restoration
	McGurk effect
	F2'
	Dual perception
	Priming effect
	Haas effect
Language-related issues	Vowel differences among languages
	VOTs among languages
	Japanese (special moras)
	L1/Ls acquisition
	Tones
	Rhythms and accents

SAMPLE TOPICS IN THE APD

In this section, some examples of these topics are described.

Lung models

Figure 1 (a) shows physical models of the human lung we developed and have used in workshops for children as well as in classes for undergraduate and graduate students (e.g., [5]). The models visualize how we breathe and phonate, and help learners to understand the whole mechanism of vowel production. We have recently adopted an anatomical animation model of the lungs as shown in Fig. 1 (b).

FIGURE 1. Lung models: (a) physical model (left) and (b) anatomical animation model (right).

Vowels in F1-F2 Chart

In this demonstration, vowels are heard when a mouse button is pressed and the mouse cursor is placed at a point on the chart of the first formant (F1) and the second formant (F2). When dragging the mouse, vowel changes will be heard along the movement. This movement can be associated with the vocal-tract shapes and the vowel spectrum, and such association helps learners to understand the acoustics of vowels systematically.

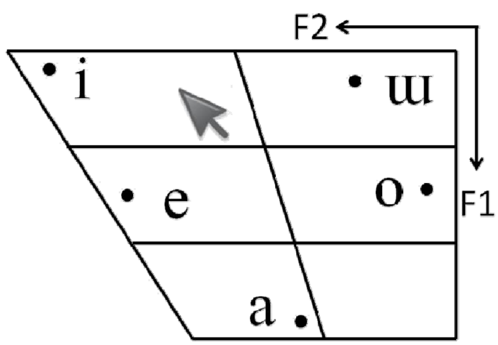

FIGURE 2. F1-F2 chart.

Vocal-tract Models

We have developed a series of vocal-tract models for different purposes as shown in Fig. 3 [5-8]. Combining any of these models with a sound source enables us to demonstrate 1) the source-filter theory of vowel production [e.g., 16], and 2) the relationship between the vocal-tract shape and vowel quality. The online demonstration contains video clips with sounds in addition to a simple description. If physical models are available the benefit to students is even greater but the video clips alone are still very helpful.

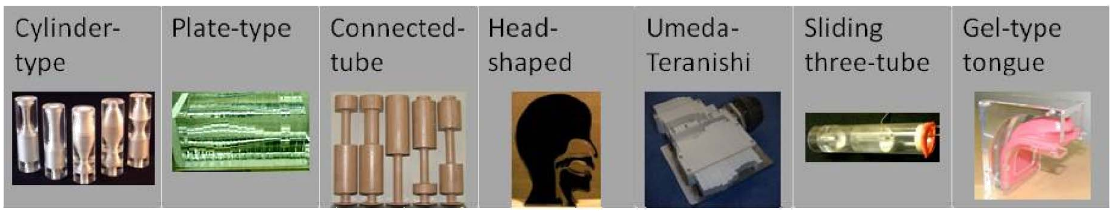

FIGURE 3. A subset of the vocal-tract models.

Digital Pattern Playback

The PP tool has contributed greatly to the development of research in speech science [13-15]. By converting a spectrogram into sound, it becomes possible to test which acoustic cues projected on the sound spectrogram are important for speech perception. One can simplify the cues and/or systematically change aspects of them, redraw a spectrographic representation, and synthesize stimulus sounds. Figure 4 shows an example of the demonstrations in the APD. The real-time version of the DPP is also effective for hands-on activities as discussed in [9, 12].

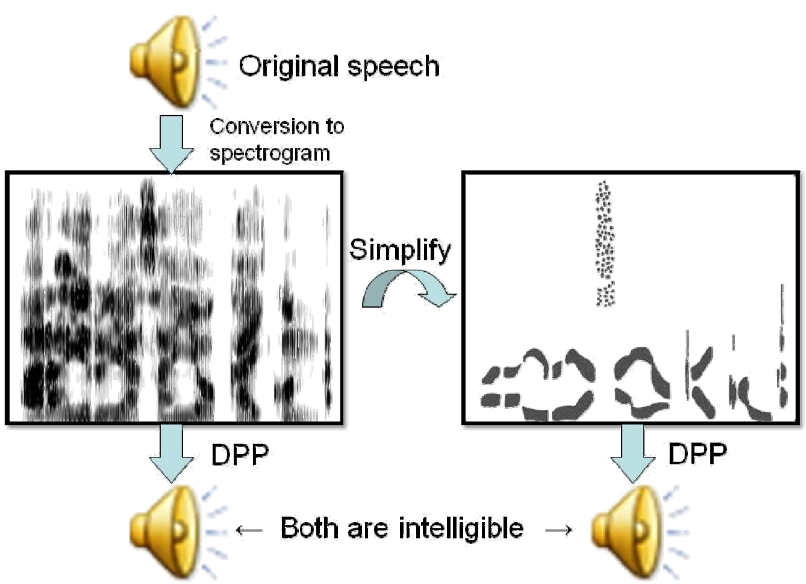

FIGURE 4. The original speech is first converted to a spectrogram representation, after which the DPP converts it back to speech sounds. At the same time, a simplified version of the spectrogram can also be converted to speech. Surprisingly, both converted speech signals are intelligible.

Stops: Voice Onset Time

In stop consonants, we can demonstrate from voiced version through unvoiced version by changing either the lag of the onset of voicing from the burst, or the voice onset time (VOT, e.g., [17]). Figure 5 shows a series of spectrograms of a consonant-vowel (CV) syllable with different VOTs. In this case, the consonant is an alveolar stop and the vowel is /a/. When listening to this continuum, /da/ is heard when the VOT is short and /ta/ is heard when the VOT is long.

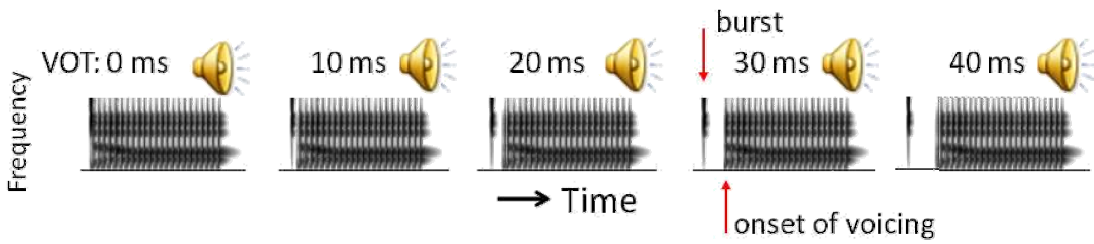

FIGURE 5. Speech sounds with different VOTs synthesized by a formant synthesizer.

Stops: F2-onset Frequency and Place of Articulation

Again in stop consonants, we can demonstrate different places of articulation by changing the onset frequency of the F2 from low to high. Figure 6 shows a series of schematic representations of formant transitions of the vowel /a/. Interestingly, CV syllables with (voiced) stop consonants can be heard with no stop bursts. The CV syllables /ba/, /da/, and /ga/ can be heard when the F2 onsets are respectively low, middle, and high (locus theory, e.g. [18]).

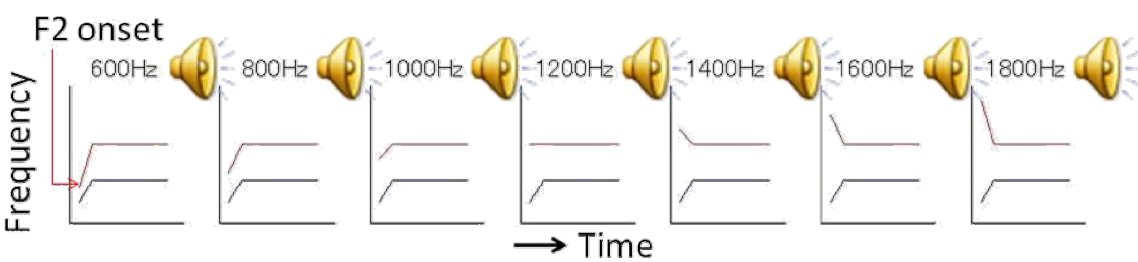

FIGURE 6. The vowel /a/ with different formant transitions. The F2-onset frequencies are varied from 600 to 1800 Hz.

Approximants and Formant Transitions

With formant transitions, the vowel in Fig. 6 sounds like a CV syllable with a stop consonant. This is true when the transitions are fast enough. When they are slow, it sounds instead like a CV syllable with an approximant. Finally, it ends up sounding like a sequence of two vowels. In the case shown in Fig. 7, /bi/ is heard when the formant transition lasts 40-60 ms, /ui/ is heard when it lasts 100 ms, and /wi/ is heard when it lasts between 60 and 100 ms [19].

Figure 8 shows a /ra/-/la/ continuum with formant transitions. As the onset and inflection frequencies increase, the perception changes from /ra/ to /la/ [20].

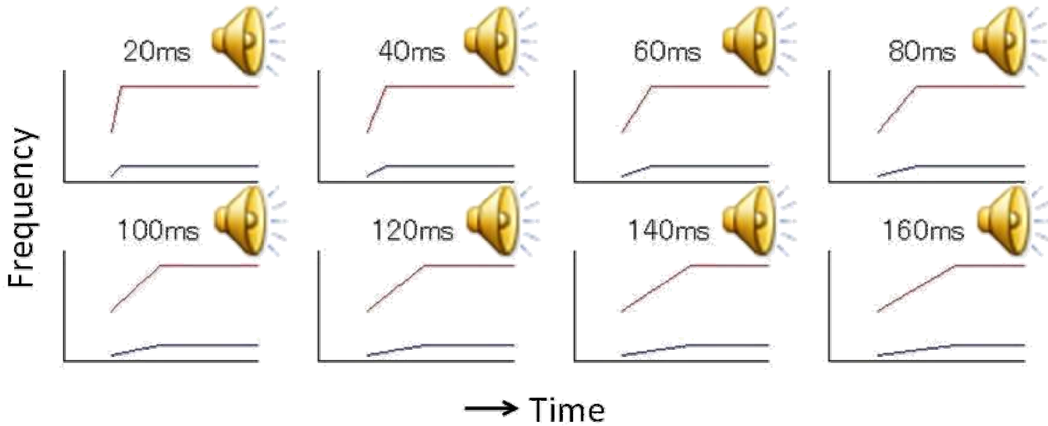

FIGURE 7. The vowel /a/ with different formant transitions. The duration of the transitions was varied from 20 to 160 ms. The CV syllables /bi/, /wi/, and /ui/ are heard when the duration is respectively short, middle, and long.

Onset freq. [Hz] at 100 ms	Inflection freq. [Hz] at 135 ms	
1609	1717	
1756	1846	
1885	1971	
2013	2085	
2150	2211	
2292	2325	
2428	2465	
2557	2573	
2688	2707	
2827	2827	

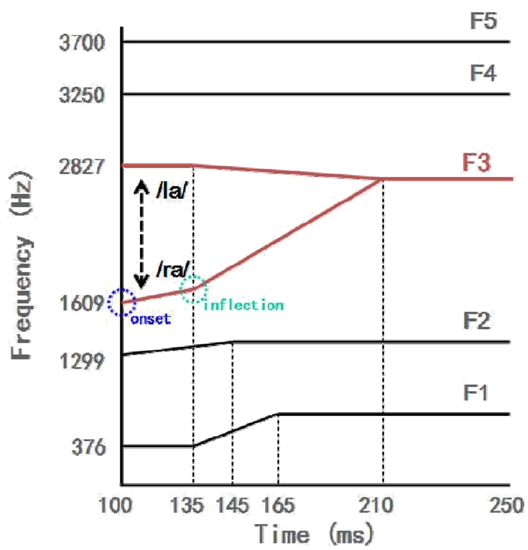

FIGURE 8. A /ra/-/la/ continuum with different formant transitions. The onset and inflection frequencies were varied as listed. As the onset and inflection frequencies increase, the perception changes from /ra/ to /la/.

SUMMARY

This paper has mainly described online demonstrations we have developed for the field of acoustic phonetics and related areas. The demonstrations are still being developed and more work is needed to further improve them. However, they have already been used in several educational situations and positive feedback on them has been obtained from users. Combining physical models and offline but real-time systems enables the demonstrations to strongly help learners by listening, seeing, and touching.

ACKNOWLEDGMENTS

This work was partially supported by a Grant-in-Aid for Scientific Research (No. 24501063) from the Japan Society for the Promotion of Science. The author thanks Kanae Amino, Keiichi Yasu, Yoko Saikachi, Hinako Masuda, Kanako Tomaru, Tatsunori Endoh, Takehiro Sannohe, Tatsuya Iwanami, and Takashi Arai for their assistance.

REFERENCES

1. http://asa.aip.org/discs.html
2. T. Arai and K. Amino, "Proposing acoustic-phonetics demonstrations," Proc. Spring Meet. Acoust. Soc. Jpn., 1479-1482, (2010).
3. T. Arai, "Learning acoustic phonetics by listening to sounds," Trans. Education in Acoustics., The Acoustical Society of Japan, EDU2011-07, 33-38 (2011).
4. http://www.splab.net/acoustic-phonetics_demonstrations/index-e.html
5. T. Arai, "Education system in acoustics of speech production using physical models of the human vocal tract," Acoust. Sci. Tech. **28**, 190-201 (2007).
6. T. Arai, "Education in acoustics and speech science using vocal-tract models," J. Acoust. Soc. Am. **131**, 2444-2454 (2012).
7. T. Arai, "Gel-type tongue for a physical model of the human vocal tract as an educational tool in acoustics of speech production," Acoust. Sci. Tech. **29**, 188-190 (2008).
8. T. Arai, "Physical models of the human vocal tract with gel-type material," Proc. of Interspeech, 2651-2654 (2008).
9. T. Arai, K. Amino, M. Sonu, K. Yasu, T. Igeta, K. Tomaru and M. Kasuya, "Hands-on speech science exhibition for children at a science museum," Proc. of WOCCI (2012).
10. T. Arai, K. Yasu and T. Goto, "Digital pattern playback," Proc. Autumn Meet. Acoust. Soc. Jpn., 429-430 (2005).
11. T. Arai, K. Yasu and T. Goto, "Digital pattern playback: Converting spectrograms to sound for educational purposes," Acoust. Sci. Tech., 27(6), 393-395 (2006).
12. T. Arai, "Digital pattern playback for education in digital signal processing and speech science," Proc. of IEEE ICASSP, 2769-2772, Kyoto (2012).
13. F. S. Cooper, A. M. Liberman and J. M. Borst, "The interconversion of audible and visible patterns as a basis for research in the perception of speech," PNAS, 37, 318-325 (1951).
14. F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman, "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am., 24(6), 597-606 (1952).
15. J. M. Borst, "The use of spectrograms for speech analysis and synthesis," J. Audio Eng. Soc., 4, 14-23 (1956).
16. G. Fant, *Theory of Speech Production*, Mouton, The Hague, Netherlands (1960).
17. L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," Word, 20, 384-422 (1964).
18. P. Delattre, A. M. Liberman and F. S. Cooper, "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am., 27, 769-774 (1955).
19. A. M. Liberman, P. C. Delattre, L. J. Gerstman and F. S. Cooper, "Tempo of frequency change as a cue for distinguishing classes of speech sounds," J. Experimental Psychology, 52, 127-137 (1956).
20. L. Polka and W. Strange, "Perceptual equivalence of acoustic cues that differentiate /r/ and /l/," J. Acoust. Soc. Am., 78, 1187-1206 (1985).