Using Arai's vocal tract models for education in Phonetics

Terri Lander† and Takayuki Arai‡
† Visiting Scholar Sophia University, Tokyo, Japan
‡ Associate Professor Sophia University, Tokyo, Japan
E-mail: terrilander@msn.com, arai@sophia.ac.jp

ABSTRACT

Dr. Arai's vocal tract models are used for instruction in a three-day Acoustic Phonetics short course taught at Northwest Christian College in Eugene, Oregon, as a part of the year 2002 Oregon Summer Institute of Linguistics. Arai's vocal tract models, designed as educational tools, are found to help students grasp basic concepts in acoustic phonetics, particularly source filter theory. A perceptual experiment investigates human perception of the output of the models at a fine phonetic level. A questionnaire reveals that students used the models to improve vowel discrimination during the perceptual experiment, and a written exam reveals students' solid understanding of the interaction between source and filter, both of which are fundamental concepts in source filter theory. Finally, we examine the students' label selections for the perceptual study and discuss implications for speech annotation.

1. INTRODUCTION

Physical models of the vocal tract have proven useful for teaching abstract acoustic concepts to students with or without prior acoustics background. Specifically, models developed by Dr. Arai based on Chiba and Kajiyama's 1941 measurements of the vocal tract [1,2] have been successful with students of differing backgrounds and ages, such as [1,3], concerning college level speech science/engineering students, and [4], with high school students. The current study incorporates undergraduate phonetics students who have no engineering experience, and again confirms that students are able to grasp abstract acoustic concepts with ease when they have access to Arai's hands-on models.

The facility with which students understand, in particular, source filter theory, is apparently due in part to the model's design. There are two types of models, both made of a transparent resin material. The first model developed consists of a series of plates having holes of varying diameters cut in the center of each plate. This is referred to as the plate-type model. When placed adjacent to one another, the holes in the plates form a tube, intended to mimic the shape of the vocal tract according to Chiba and Kajiyama's measurements of the vocal tract for the Japanese vowels [i,e,ɑ,o,ɯ] [1,2].

The plate model has been found to be particularly effective in the classroom, because students are able to manipulate the shape of the tube themselves and see, in real-time, how tube shape relates to acoustic output. Plate configuration for each of the five Japanese vowels is specified. Besides allowing students to see how resonator shape effects acoustic output, another strength of the plate model is its potential to create more vowels than just the five, with a simple rearrangement of the plates.

A second type of model is the one-piece, cylindrical model, which is also based on Chiba and Kajiyama's measurements. The benefit of these models (there is one for each of the five Japanese vowels) is that the model is in one piece and in that sense is intuitively more similar to the vocal tract. The cylindrical model is also easier to set up, making it suitable when giving a series of auditory examples during a lecture. Formal descriptions of the models are given in [1].

The goal of the paper is to investigate whether and how Arai's vocal tract models are effective for helping to convey abstract concepts of acoustic theory to Phonetics students who have little or no acoustics background. To do this we:

- Conduct and discuss a perceptual study that examines human perception at a more fine phonetic level than has been done previously, [1], highlighting patterned variations from the expected response in the annotation of the perceptual experiment.
- Examine questionnaire responses and discuss how students used the acoustic models for vowel discrimination and the implications for the usefulness of the models.
- Discuss how important the models are in helping students understand speech theory, as evidenced on written exam and informal student comment.
- Posit implications of this study for speech annotation.

2. METHODOLOGY

The acoustic phonetics short course discussed here comprised the final three days of a summer-long articulatory phonetics class. There were 29 students in the class. 23 of the 29 students had no prior acoustics experience or training. Four students had some exposure

ISBN 1-876346-48-5

to physics in high school. Two students had one college course in physics prior to this acoustics seminar. One student reported that he hears better in his left ear than in his right ear (however, this did not seem to affect his performance). The course consisted of approximately 5 hours of lecture and lab and 1 hour for a written exam. During the lecture the acoustic models were alluded to frequently. Students were given a good deal of time to experiment with the plate and cylinder models, as well as the electro larynx and whistle-type larynx included in the set of manipulatives produced by Dr. Arai. Students were also given problems to solve, such as making the models for unrounded vowels sound more like rounded vowels. They did this by manipulating the plates in the plate-type model.

29 students took part in a perceptual experiment. To practice, five tokens were given by piping sound through the cylindrical models. The answers were discussed. The actual perceptual test consisted of five experimental tokens as well, but we used the five plate arrangements based on Chiba and Kajiyama's measurements for the Japanese vowels [i,e,ɑ,o] and [ɯ]. Students were not told which vowels to expect, and were not allowed to discuss their answers with anyone during the experiment. The tokens were administered to the class as a group. The experimenter replayed the sounds as requested by students. Participants were instructed to select the IPA symbol (no diphthongs allowed) they felt most closely resembled the sound produced by the model.

3. RESULTS

	ɨ i ɪ y (i)	e ɛ ø (e)	a ɑ (ɑ)	o ɤ ɔ (o)	ɯ (ɯ)	ʌ ə
i	**27**	1				
e	1	**24**			1	
ɑ			**28**			1
o		1		**19**		7
ɯ	1	3		2	**5**	12

Table 1: Results of the perceptual study involving 29 students discriminating the 5 Japanese vowels intended by the plate-type model (i,e,ɑ,o,ɯ). To highlight patterned variation, non-contrastive, phonetically close vowels selected by students are merged, as shown in the top row. Vowels selected less than three times as responses on the perceptual test are not shown in the table (this includes œ, ʏ, ỹ , ɵ, u, æ and ʊ).

i	**e**	ɑ	**o**	ɯ
27/29	24/29	28/29	19/29	5/29
93%	82%	99%	65%	17%

Table 2: Shows "percent correct" identification for each vowel, where "correct" means the students chose vowels similar to those intended by the model, and where similar is defined in the first row of Table 1. Specifically, the ratio expresses the number of times the students identified the vowel they heard as (or close to) the sound the model was intended to produce, out of a total of 29 instances. Secondly, each ratio is expressed as a percentage. The idea of a "correct" identification is addressed further in the footnote on page 4.

	Round	Unround	%Unexpected
i	**10**	19	34% 10/29
e	**9**	20	31% 9/29
ɑ	**0**	29	0% 0/29
o	16	**13**	44% 13/29
ɯ	10	**19**	65% 19/29

Table 3: Displays the number of times the 5 experimental tokens produced by the model were identified as round or unround, by nature of student label selection. Numbers highlighted in gray indicate "unexpected" identification of the feature rounding, that is, students chose IPA symbols which by definition imply the feature rounding, when the model actually intended an unround vowel (according to Chiba and Kajiyama's measurements) and vice versa. The far right column indicates the percentage of "unexpected" identity, according to this definition.

Questionnaire responses referencing the models

Students had the following responses when asked to list the strategies they used to take the perceptual test:

"Looked at shape of chamber especially where the sound comes out last—would indicate lip rounding and vowel height"

"I looked at the shape of the tube and listened to hear what vowel it was."

"Look at shape of 'mouth'" (*referring to model opening*)

"Looked at shape of tube"

"Listened and tried to guess from looking at the tube…"

"Looked at the shape of the resonators and was able to guess either rounded or unrounded, etc." ("*resonators" is a loose reference to the vocal tract model*)

"I looked at the tube shape and listened to the sound."
"Careful listening, or I watched the shape of the one-piece sounds" (*referring to cylindrical models in practice set*)
"I looked at the shape a little, but mainly tried to put it in an English word."

Table 4: Quotations from students that specifically refer to the acoustic models and how they found the models useful during the perceptual experiment.

4. DISCUSSION

Perceptual Experiment

When students chose from all IPA vowels symbols, there was more variability than in past perceptual experiments where only the five Japanese labels were options [1]. For instance, in the current study, only four out of 29 students identified all five vowels as the Japanese vowel the model was intended to produce. However, Table 1 shows that variability mainly concerned acoustically similar vowels, suggesting the models produce vowels recognizable as the intended vowels (except for [ɯ], see below). It may be helpful to point out that Arai's models are based on Chiba and Kajiyama's measurements, and they may not accurately produce the five Japanese vowels. Part of the motivation for this perceptual experiment was to see how close to the intended vowels people perceived the output of the models to be.

Table 1 displays a confusion matrix of label assignments made during the perceptual experiment. The results underscore patterned variability. Specifically, much of the variability in label assignment occurred between vowels that are close in quality and also non-distinctive in Japanese. Therefore, to present the results, certain phonetically close and non-distinctive vowel labels were merged, as shown in the first row of Table 1. The symbol in parenthesis in the second line indicates which model the merged vowels concern, or rather, which vowel the model was intended to produce.

Interestingly, none of the vowels except for [ɯ] was ever identified as another of the 5 vowels; it happened three times with [ɯ]. Outliers account for 11answers, 5 of which concern the difficult to identify vowel [ɯ].

Table 2 summarizes results from Table 1, indicating 93% correct identification for [i], 82% for [e], 99% for [ɑ], 65% for [o], and 17% for [ɯ]. The model for [ɑ] is the most reliably identified, followed closely by [i] and then less closely by [e]. Identification of the output of the model for [ɯ] had a good deal of unpatterned variation in that it was mistaken for every other vowel except [ɑ]. This could be attributable in part to the fact that the back unrounded vowel is not native to any of the students, but more likely, it suggests the model does not produce the vowel [ɯ] as distinctly as it does the other vowels. The model for [o] was also less reliably identified, but not to the same degree. One may argue that since the monophthongs [e] and [o] do not occur in English this could have affected perception, but as students were specifically instructed to use monophthong labels, residual effect from phonological expectation is doubtful.

Table 3 was created because it was noted that a high percentage of unexpected identification (see caption, Table 3) was correlated with the feature rounding. We see that 34%, 65%, 31% and 44% of unexpected identifications of [i,e, o,ɯ] respectively are associated with labels that by definition contain the notion of rounding. The rounding issue is interesting, and it highlights an area that a future generation of model, a generation not based solely on Chiba and Kajiyama's measurements, could address. Furthermore, whether or not the tools perfectly model an abstract notion of "expected" vowel (or feature), does not diminish the usefulness of the tools as teaching aids, because the relationship between tube shape and acoustic output is unrelated to the issue of whether or not the models are able to produce any particular vowel.

Questionnaire

Table 4 confirms that students understand that the shape of the tube is related to acoustic output, a crucial concept in source filter theory. Specifically, in Table 4 we see that nine of the 29 students said they looked at the shape of the model during the perceptual experiment. The transparent design of the model helped them to use the model for vowel discrimination. Two students said they examined the end of the tube to determine rounding on the vowel.

Importance of acoustic models

The questionnaires just discussed, informal student comments, videotaped student laboratory scenes, as well as short answers given on a written exam all suggest that the models greatly facilitated students' understanding of source filter theory. The students were excited about the models.

Most students were able to define source filter theory in their own words in a written exam, especially the concept that a source sound representing all frequencies at equal amplitudes is modified (filtered) by the vocal tract and the result is speech sounds in which most frequencies are dampened but those closest to the natural resonating frequency of the filter, in its particular configuration, are amplified. In the five times I've taught this short course (without the use of the models), I have never seen a group of students grasp that concept so thoroughly.

Finally, access to the electro larynx was helpful. First, students heard the buzz sound produced by the artificial larynx, and were told this is like the sound produced at the larynx, or sound source. Second, students realized how the vocal tract filters glottal sounds when they mouthed

ISBN 1-876346-48-5

words while placing the electro larynx at their larynx. They saw that the electro larynx emitted speech-like sounds in the absence of pulmonic force, providing clear, intuitive separation between the glottal source and vocal cavity acting as a filter.

In summary, although it could be argued that students with training in articulatory phonetics already have an intuitive sense of the importance of filter shape on acoustic output, so that the success realized in this study may be due in part to prior knowledge, we do not feel this is entirely the case. Given the reliance students had on the models during the perceptual experiment (Table 4), as well as answers on the written exam, a strong case can be made that the vocal tract models coupled with the electro larynx confirmed phonetics students' already developing intuitions regarding the filter (vocal tract), and further enlightened students as to the precise relationship between the source (glottis) and the filter in source filter theory.

Implications for speech annotation

Interestingly, the four students who labeled all five vowels correctly used a "phonemic" strategy. When queried, they said they either assumed the vowels were the same as the five in the practice session (a correct assumption) or they tried to map the vowel they heard to a n English or Cardinal vowel. The 25 students using non-phonemic strategies did not label all of the vowels with the intended label, though they usually chose a vowel of similar quality. This supports motivations for a methodology described in [9] designed for normal hearing and hearing impaired speech, which crucially relies upon phonemic motivations (by which is meant mapping a sound to the closest phoneme) for label choices. Lander (2000) contends that labeling produced with such strategies is not only more consistent between labelers, but it is also *better* because labelers more closely identify the actual sound produced, as determined by a consensus of labelers.[1]

[1] Whether or not sound identification is possible, or whether there exists an actual identity of a sound, is an interesting question. It is the first author's contention that identification is possible to approximate in most cases, as evidenced by the fact that in the face of a great deal of phonetic variability, speakers of a language generally understand each other. Also, experience shows that the identity of a sound can be established most reliably when sounds are identified within a phonemic framework shared by multiple labelers. If "phoneme" is defined as a meaningful sound unit, it can be said that when sounds appear in meaningful contexts, identification is possible. At this point one might argue that sounds uttered in isolation, such as during a perceptual experiment, have no meaning, and therefore cannot be identified. However, if one approaches sound identification and label assignment in terms of mapping a sound (whether isolated or otherwise) to the closest recognizable phoneme (which is not the same as the phonologically expected phoneme), and if a uniform label set such as the IPA is chosen, consensus for sound identification is possible, because an artificial meaningful context is thus imposed on the data. Several pieces of evidence argue in favor of this methodology: 1) lessons learned from lower inter-labeler agreement in less phonemically based labeling as was done in [10,11], 2) unpublished research using [9] for labeling hearing impaired and hearing speech and 3) informal observations of labelers by the first author over the course of a decade working through these issues. These points conspire to suggest that to the extent "true" identification of sounds is possible, it is most reliably accomplished when 1) label selection is phonemically based, 2) choices for label assignment are minimized according to phonemic rather than phonetic criteria, and 2) listeners select from a uniform set of labels.

5. CONCLUSION

In the current study we saw that acquisition of abstract speech theory concepts was crucially linked to availability of the acoustic models of the vocal tract developed by Arai. Responses on a questionnaire, casual comments made by students, exam answers, and results from a perceptual study confirm the usefulness of Arai's hands-on models in speech education for phonetics students.

REFERENCES

[1] T. Arai, "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity," J. Phonetic Soc. Jpn., 5(2):31-38, Aug. 2001.

[2] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, 1941.

[3] T. Arai, N. Usuki and Y. Murahara, "Prototype of a vocal-tract model for vowel production designed for education in speech science," Proc. of Eurospeech, 4:2791-2794, Aalborg, Sep. 2001.

[4] E. Maeda, T. Arai, N. Saika and Y. Murahara, "Lab eperiment using physical models of the human vocal tract for high-school students," Proc. of the First Pan-American/Iberian Meeting on Acoustics, Cancun, Dec 2002.

[5] T. Arai, "An effective method for education in acoustics and speech science: Integrating textbooks, computer simulation and physical models," Proc. of the Forum Acusticum Sevilla, Sep., 2002.

[6] N. Saika, E. Maeda, N. Usuki, T. Arai and Y. Murahara, "Developing mechanical models of the human vocal tract for education in speech science," Proc. of the Forum Acusticum Sevilla, Sep., 2002.

[7] E. Maeda, N. Usuki, T. Arai and Y. Murahara, "The importance of physical models of the human vocal tract for education in acoustics in the digital era," Proc. of China-Japan Joint Conf. on Acoust., 163-166, Nanjing, Nov, 2002.

[8] T. Arai, E. Maeda, N. Saika and Y. Murahara, "Physical models of the human vocal tract as tools for education in acoustics," Proc. of the First Pan-American/Iberian Meeting on Acoustics, Cancun, Dec 2002.

[9] T. Lander, "A Labeling Methodology for Hearing and Hearing Impaired Speech," Technical Report, Center for Spoken Language Research, University of Colorado, 2000.

[10] T. Lander, B. Oshika, R. A. Cole, and M. Fanty, "Multi-language Speech Database: Creation and Phonetic Labeling Agreement" Proc. of ICPhS, Stockholm, 1995.

[11] R. Cole, B. T. Oshika, M. Noel, T. Lander, and M. Fanty, "Labeler Agreement in Phonetic Labeling of Continuous Speech", Proc. of ICSLP, Yokohama, 1993.