TEACHING ACOUSTIC PHONETICS

J. Vaissière
CNRS-URA 1027, Institut de Phonétique,
19 rue de Bernardins, 75005 Paris, France

1. INTRODUCTION

Ladefoged [1] stated in the introduction to this symposium that we should think of a good phonetician as someone who has a good grasp of phonetic principles and understands the issues in speech production, perception and acoustics.

Since the acoustic signal is the central element in the speech communication chain, a rather deep knowledge of speech sounds, how to characterise them acoustically, how such characteristics can be produced in an human vocal tract, how they are perceived, plays a dominant role in a unified understanding of speech. Acoustic phonetic knowledge permits an understanding of the interdependence of issues in speech production, perception and acoustics. A solid understanding of the complexity of human speech relies on appropriate knowledge of the acoustic phonetics field.

My talk is addressed to teachers faced with the problem of teaching acoustic phonetics in very large class rooms. At my institution, undergraduate linguistics students obligatorily attend year-long phonetics courses three hours a week for the first two years and a minimum of 1 1/2 hour during the third year. The degree of motivation for studying phonetics is not uniform among the students. It is important that they must all accept that they have to learn a series of notions that they may perceive at first as too technical.

The talk is also addressed to teachers of phonetics who are seeking ideas for how to transmit in the most efficient possible way the necessary background to students and professionals from various disciplines (linguists, engineers, professors from foreign universities, medical doctors, physiologists, computer scientists, speech pathologists, etc.). Those students are motivated, they have access to phonetic laboratory facilities, and they will soon be engaging in multidisciplinary research involving phonetics. Intensive courses in acoustic phonetics allow them to find phonetic problems for themselves and to solve the problems using a limited mathematical background (see for example, a paper on Arabic fricatives [2] and on the singer's formant [3], presented in this congress). This sort of acoustic knowledge is like mathematics for an engineer. The demand of intensive phonetics courses is expanding globally.

2. WHAT TO TEACH?

All students should be given the necessary background to understand the links between

- **basic acoustic laws** underlying generation and propagation of sounds,
- **articulatory manoeuvres** for producing the speech sounds in an human vocal tract,
- **acoustic characteristics** of the resulting sounds, and
- the interpretation of such acoustic characteristics by the **perceptual mechanism.**

In other words,

1) Students should be given the background to understand the characteristics of the spectrographic representation of speech in relation to articulation.

2) To do so, students must understand how speech sounds are created by talkers.

3) So, students should know about the control of airflow (sound source generation) and of vocal tract shapes (acoustic characteristics of speech sounds)

4) Consequently, they need to understand the relationship between the vocal tract shapes (VT) and the acoustics (Fn) without or with only simple mathematics (computer programs can be helpful for the acoustic calculations).

To summarise, there are five basic topics:

1) - what is a **wave**?

Periodic, *quasi-periodic* and *non-periodic* waves (noise).

A *sinusoidal* wave is a periodic wave having a single frequency.

Periodic waves (vowels, nasals, etc.) are composed of '*harmonics*'. They are the sum of sinusoidal waves, with their frequencies n * Fo, where n = 1, 2, 3, ..., n, and Fo is the fundamental frequency in Hz.

Vowels are quasi-periodic and *fricatives* non-periodic.

2) - What is **resonance**?

A resonance manifests itself as a *damped oscillation* of a wave . It is characterised by its *natural frequency* and *damping*.

The *vocal tract* acts as a set of resonators. The resonance frequencies depend on the shape (area function) of the vocal tract.

3) **Simple acoustic tube**, and their relation to resonances:

Resonance modes like the *quarter wave length* resonances, *half wave length* resonances, and *Helmholtz* resonance.

4) Vocal tract acoustics calculations using **connected simple tubes** representing vowels.

Comparison of acoustic and perceptual calculations and that using a simulation program.

In order to obtain the "simple tubes" representation of the vocal tract, X-ray pictures (or MRIs) in the midsaggital plane are used.

5) About **sources:**

Vocal-fold vibration (measurements using the glottograms, EGG) etc.

Aerodynamics principles underlying noise generation.

3. HOW TO TEACH

In the following text are listed a set of traditional and less traditional "tools" for teaching acoustic phonetics.

a) Tuning fork, pendulum and anecdotes

Most of the available introductory books contain valuable information on how the courses can start by using traditional tools.

Sounds are acoustic vibrations propagated to the ear through air. What is a *vibration*, and what *propagation* means? A **pendulum**, such as the one that I will ask Mary Beckman to set up gently into motion, is a mass attached to a string pulled into oscillation. A simple watch or chronometer may be profitably used to introduce in front of the students the notions of *movement, amplitude* of movement, *periodicity*, influence of the length of the string on periodicity, *cycle, frequency, sinusoidal* oscillations, *damped* oscillations, representation of the displacement on a *graph* (displacement versus time, the magnitude of displacement versus frequency or resonance curve), etc.

Tuning forks allow students to *hear* sounds differing by *pitch* only, or by *loudness*. Two identical tuning forks can be used in a very effective manner for an introduction of the notion of *natural frequency* and *resonance*, as I will ask Peter Ladefoged to demonstrate in front of you. All demonstrations with tuning forks can be even made more fruitful by systematically providing the students with *waveforms, spectral slices* and *spectrograms* of the sounds.

A useful way to attract the attention of the class is to start a section with one of the traditional **anecdotes** (the resonant Tocama Narrows Bridge [4], the glass broken by the voice of a singer, etc.) and to let them find by themselves various examples of resonances in daily life: child swing, grand-father's clock, undesirable car vibrations at a certain speed, etc..

Peter Ladefoged's **book** 'Elements of acoustic phonetics' first published in 1962 and this year to be re-issued in a revised version with additional material on digital speech processing [5], offers a coherent 'parcours pédagogique" for teaching the basic notions at the most elementary stage.

b) Spectrogram decoding as a basic complementary course

The best way I found to prepare the students for a unified understanding of speech is extensive practice in **spectrogram decoding.** At the very beginning, the students learn to characterise non speech, familiar sounds, and synthesised sounds, with varying intensity, pitch, formant frequencies. Then they are taught how to recognise the five vowels /oe/, /i/, /a/, /u/, /y/ and nasal /a/ in connected speech. The perturbation theory is then used to introduce the influence of the consonants on the neutral vowel [oe], and the formant directions in the transitions in /peop/n, /toet/, /koek/ and /roer/, etc. At the end of the second year of progressive learning, most of the students become decent spectrogram readers. The student is taught to address, at the very beginning, the same detailed attention to how each French sound is produced, its pertinent acoustic characteristics and its corresponding perceptual sensations, using a speech editor to extract the sound from connected speech. Needless to say, teaching the basics of Fant's theory is facilitated when addressing fairly good spectrogram readers.

c) Easy access to speech analysis facilities for the teacher and the student.

Computer speech analysis allows phonetics teachers to do a quick printing of high quality spectrographic representations, waveforms and spectra, which are distributed to the undergraduate students who don't have direct access to speech analysis facilities.

The availability of an inexpensive analysis software (such as SNORRI) [6] producing very high quality spectrograms and affordable Audio Cards may allow in the near future to equip a number of PCs outside of our laboratory, giving to some undergraduate students access to speech analysis facilities.

e) Speech analysis programs as pedagogical tools

Demonstrations with analog Sonagraph, pitch detector and filter banks have a strong pedagogical impact on students, and should be used whenever possible in parallel. Video recordings of the demonstrations using the analog devices can be displayed in the classroom.

Since analysis of the speech signals employs digital techniques, the acquisition of knowledge of digital signal processing, in particular **filtering** and **sampling**, has become necessary for all students. One way of teaching is to let them perform first *inadequate* filtering and sampling, and then to teach the basic knowledge necessary to do the right thing.

e) Pencil, paper and eraser for the students

Needless to say, a very important part of the teaching should be done with pencil, paper and eraser:

- hand measurement of the fundamental frequency directly from the signal in the time domain, or using the tenth harmonics in the frequency domain,
- estimation of the area function from a few typical VT sagittal profiles,
- modelling of simplified VT configurations corresponding to [i], [a], [u] and [y] by a small number of connected tubes, hand calculation of the natural resonances of each isolated tube, identifying the resonance mode, and determination of the associations between resonance cavities and formants.
- hand calculations of natural frequencies in coupled resonators representing different place of constriction along the vocal tract for consonants,
- etc.

A VT acoustics simulation should be used only after the students have a good understanding of what is going on, for example,

- listening to the sounds corresponding to the hand calculations they have done,
- estimating the effect of radiation impedance and yielding walls, and
- comparing modelling with simple tubes with a more realistic model of VT shapes,
-etc.

Too early use of computer simulation (a source of fun, indeed) may hinder correct understanding.

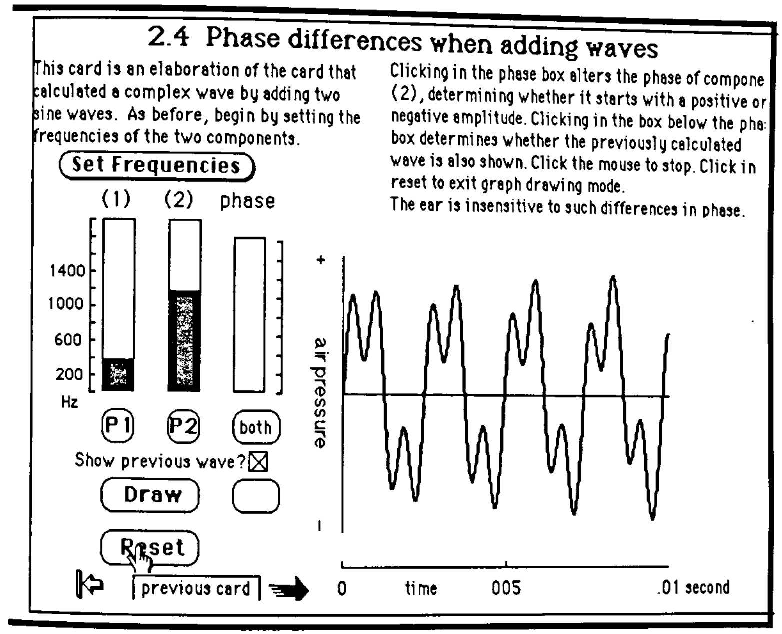

Figure 1: One of the sHyperCard tsack "Acoustic Phonetics", illustrating the effects of adding waves and of phase differences. It is possible to set the frequencies of the two components, sine wave P1 and P2, to listen to each component separately, and to observe the complex wanves and to listen to the resulting sound. It is then possible to alter the phase of component P2 in relation to P1, and to compare the resulting sound with the previous sound.

f) "Acoustics Phonetics" HyperCard teaching stack for the instructor and the students

Not all the necessary notions are easily taught to beginners. Let us take some examples.

A critical point is to let the undergraduate students understand how any waveform can be analysed as the sum of two or a greater number of sine waves (*Fourier analysis*). One way is to let them add pure tones (sine waves) on graph paper to draw a complex wave shape, and then to propose to the students the reverse problem: the decomposition by hand of a complex wave into its two sine components.

I have recently tested with success one of the lesson in 'Acoustic Phonetics' HyperCard stack [7]. The card allows the learners to hear two sine waves separately and the corresponding sounds (see Figure 1). A short demonstration will be done in front of you by the author of the program, Peter Ladefoged: selecting particular frequency for the first sine wave P1, hearing P1, choosing particular frequency for the second wave P2, listening to P2, combining both P1 and P2, observing the resulting complex waveform and listening to it.

This kind of demonstration seems to be very effective as a first approach to Fourier analysis. Let us take a second example. Intuitively, the students assume that two waves that are very different on paper should sound very different, and have trouble accepting that the human ear is not particularly sensitive to (constant) *phase*. The previous HyperCard program allows one to change the phase relation between P1 and P2, to listen to the resulting sound, and then to compare it with previous sound.:

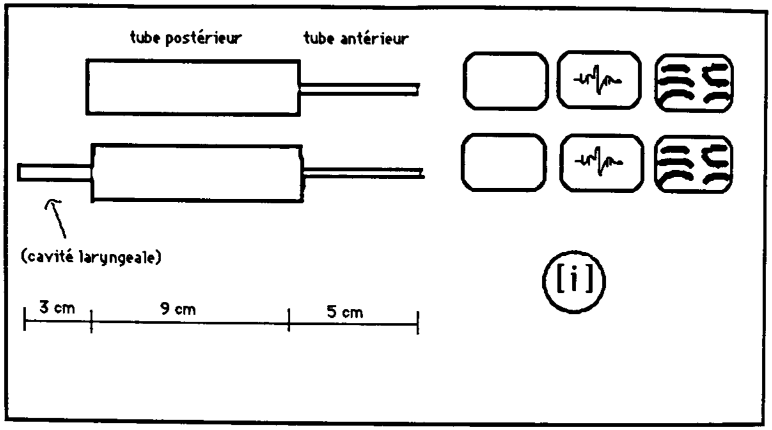

Figure 2: The vowel [i] synthesised with two (9 cm long and 12 cm2 in area, and 5 cm long and 1 cm2 in area) or three tubes. By clicking on the appropriate box, it is possible to listen to the synthesised sounds, or to observe the corresponding signals and spectrograms .The demonstration aims to show the improvement by adding a third tube (3 cm long, and 2 cm2 in area) simulating the laryngeal cavity.

The comparison done by the ear speaks for itself. Practical experience helps one to believe what the books say.

d) Synthesis and simulation programs

A course on acoustic phonetics should provide a good understanding of the links between (i) production and the acoustic properties of the signal, and (ii), the acoustic properties of the signal and perception.

Fortunately, the first link, i.e. the link between VT configurations and acoustic properties of the signal is rather well understood. **Fant**'s famous book "Acoustic Theory of speech production" [8] is an authoritative. guideline for the teacher of phonetics. Note that much of the mathematical specification contained in Fant's book cannot be acquired by average linguistic students (and luckily, it is not strictly necessary). Lehiste's "Readings in Acoustic Phonetics" is also an excellent source of inspiration for setting up the content of the courses [9]. Our research students are invited to use also intensively Stevens's many articles ([10] for example).

The other link, i.e. between the acoustic characteristics of the sounds and what is effectively **perceived** by the listener(s) is less well understood. Of particular interest are the acoustic explanations of
- **compensating** and **reinforcing** movements (two simultaneous changes in VT, such as lip rounding and larynx height can reduce or enhance acoustic effect,
- sound **confusions**, and
- sound **changes**.

Both links can be studied in parallel, thanks to modern computer technologies which provide quick calculations and sound facilities. Vocal tract acoustics simulation programs make it feasible to manipulate one articulatory parameter at a time (tongue position, jaw position, lip height and width, lip protrusion, etc.). Formant synthesisers make it feasible to manipulate one acoustic parameter at a time (formant frequency, bandwidth, etc.) Thanks to both type of these programs, it is possible to investigate the links between VT (articulation) and Fn (acoustics) and the auditory responses (perception) altogether.

To summarise, the programs I used most frequently in acoustic phonetics courses are:

a) a version of Klatt's **formant synthesiser** [11] on PC
b) **Two-tube** model, **Fant's model**, a variable number of tubes model, with optional on/off radiation load, and optional rigid/yielding walls (programs written by Maeda, for PCs). The insertion of a noise source is under development.
c) Maeda's **vocal tract acoustics simulation** program on PC (with nasal cavities)
d) and recently "Acoustic Phonetics" HyperCard Stack from Ladefoged [7], on Macintosh.

In addition, vocal tract acoustic simulation programs developed at ICP at Grenoble [12] have been implemented on Macintosh and there is also an UCLA vocal tract model available (also on Macintosh). Note that we still lack at our institute a good simulation program on aerodynamics, and on the glottal source. Suggestions and eventual offers from the audience are really welcome.

It is important to note that the same programs are used:
- for demonstration in front of the class-room for all students,
- for a series of assignments given to the students having access to the lab facilities (some of the exercises are inspired by the excellent course initially set up by Dennis Klatt at MIT) [13] and
- as a research tool (cf [2] and [3]).

If there is enough time, I would like to show you four other demonstrations of what can be done with simulation programs.

First, any **formula** becomes attractive and easier to internalise if each of its variables can be changed at a time in front of the students. For example, concerning with the quarter wave length resonance equation, we can change the sound velocity of the air simulating speech in helium, and manipulate the length of the tube simulating the neutral vowel /oe/ produced by an infant or giant vocal tracts.

Second, what is the perceptual effect on formant patterns of a **laryngeal cavity** to the specification of a transfer function of a vowel calculated by computer simulation? This can be demonstrated by synthesizing sounds with and without the laryngeal cavity (Fig. 2). The students can observe the appearance of a supplementary formant around 2800 Hz on [i] spectrum and they can hear the differences. As you can judge by yourself, the new [i] (with added "laryngeal cavity") sounds nicer, more clear than the sound corresponding to a two tubes representation.

Third, simulation may help the students to understand the basic idea underlying most of a **theory**, which may seem, as first, rather abstract. The following 'figure parlance' serves as an introduction to *Stevens's quantal theory* of speech [10]. You are invited to listen to simulated /a/ sounds with two-tubes model and transcribe what you hear on a sheet of paper. In the critical region where the change in formant-cavity affiliation occurs, a stability in the sound quality is perceived but in the two outside regions, the same amount of displacement leads to a change in the quality of the vowels.

Fourth, the following card demonstrates the relation between the degree of velar lowering and the degree of perceived nasality for the vowels /i/, /a/ and /u/ [14]. Such demonstration can be used to draw a link with the **phonology** course on the distinctive use of nasalisation in the vocalic systems.

g) Portable computer for the teacher

In Rousselot's time, the very few students of phonetics in Paris were acquiring the necessary notions by doing experiments in the phonetic lab, under the supervision of their teacher. Thanks now to availability of portable computers, such as the one I am using now in front of you, it is possible to bring **part of the lab** into larger classrooms. A small portable computer can quickly become your most comprehensive assistant for teaching, and offer you technical assistance with rapidity, patience, precious memory, and constancy.

A portable computer can advantageously replace in many cases a **tape-recorder.** Any sound can be stored in the computer, quickly accessed, segmented using a simple speech editor such as SoundEdit on Macintosh, and repeated at will, allowing one to set up spontaneous identification and discrimination experiments in the class-room. Note that the speech files can now be easily transferred between PC and Macintosh computers and stored on disquettes eventually distributed to the students as a complement to the written version of their courses.

New technics allows also to project the display of the computer onto a large screen via a **transparency** projector. Traditional silent pictures can be advantageously replaced with transparencies with sound possibilities.

k) The use of the French vocalic systems

Finally, the use of the French vocalic system seems pedagogically well adapted for acoustic phonetics courses.

According to my own experience decoding French spectrograms is easier than decoding English spectrograms (at least in carefully spoken isolated sentences). The use of rich French and well tempered vocalic system makes it possible to illustrate distinctive labialisation and nasalisation.

Two vowels are particularly precious /y/ and nasal open back vowel /ɔ̃/. The inadequacy of an **F1 versus F2-F1** representation for the vowels is easily demonstrated to the students: /i/ and /y/ can have about the same F1 and F2, but still sound very different and are almost never confused on spectrograms. **Exchange of cavity affiliation** for a formant is also easily understood by the [i]-[y] pair. Nasal vowels are useful to illustrate the acoustic and perceptual consequences of **side cavities**. The existence of the nasal vowel [ɔ̃], which is [+nasal] and [+round] is particularly adapted for studying **co-ordinated anticipatory coarticulation** (the lips and the velum).

3. CONCLUSION

A number of good standard phonetics books, Fant's book, classical articles, traditional pendulum and tuning forks, paper, pencil and eraser, speech analysis facilities, a portable computer, the availability of pedagogical multimedia computer programs, extensive use of computer simulation, etc. seem to facilitate the task of teaching acoustic phonetics to all types of students. Furthermore the use of synthesis and vocal tract acoustic simulation programs seem indispensable for teaching the link between articulation, acoustics and speech perception. The attempt by Peter Ladefoged setting up a multimedia course on Acoustic Phonetics on HyperCard, with self-training and sound facilities seems to me a new way of teaching that should be pursued: listening is very important for phoneticians, and traditional books are too silent for allowing a deep understanding of certain aspects they describe. Transforming classical figures taken from the literature as "speaking pictures" and systematically attaching speech files to the Word version of research reports seems to me very useful.

It is important to make the best use of technological potential. The continuous technical progress, data base on CD-ROM, multimedia (text + sound + images), extended network possibilities, broadcast and interactive television, will probably change the way of teaching and of learning in the years to come. In particular, a complete platform for improving the teaching of all aspects of phonetics has become technologically feasible. It has still to be done. It may be very difficult to obtain funding for such a project. It is also difficult to find the necessary time to collect all the free data available. If a collaboration via e-mail and Interet between enthusiastic teachers of phonetics and engineers willing to share their talents and resources could be established, some progress can be made before the next congress of phonetic sciences.

But already now, and thanks to generous donors, it is possible, without any expense or with a moderate expense to improve teaching environment for the large mass of students who don't have access to laboratories facilities. It seems also possible to provide the students who want to engage in multidisciplinary research involving acoustic phonetics with the tools they need to solve the problems, using a limited mathematical background.

ACKNOWLEDGEMENTS

Thanks again to the generous donors, engineers and phoneticians alike, Shinji Maeda from CNRS, Gérard Bailly and Louis-Jean Boé from ICP in Grenoble, Peter Ladefoged at UCLA, Yves Laprie from Nancy for SNORRI, and also to Ken Stevens for his inspiring teaching. Thanks also to my students, for the fun I have to prepare courses for them.

REFERENCES

[1] Ladefoged, P. (1995), "Teaching Phonetics", Introductory comments, *ICPhS 95.*

[2] Yeou, M., & Maeda, S. (1995). "Uvular and pharyngeal consonants are approximants not fricatives: An acoustic Modelling study", *ICPhS* 95, . Stockholm.

[3] Pillot, C. (1995). "Production and perception of the singer's formant". *ICPhS 95*, . Stockholm.

[4] Spectacular Disastrous, Warner Brother's video.

[5] Ladefoged, P. (1962), *Elements in acoustic phonetics"*, University of Chicago Press, Chicago (last revised version, 1995).

[6] Laprie, Y. (1989), "SNORRI: an Interactive Tool for Speech Analysis", *Eurospeech 89* Paris, Vol. 1, pp 613-616.

[7] Ladefoged, P. *Acoustic phonetics*, HyperCard Stack.

[8] Fant, G., (19xx), *Acoustic theory of speech production,* The Hague.

[9] Lehiste, L. (1970), *Readings in Acoustic Phonetics,* MIT Press, Cambridge.

[10] Stevens, K., (1989), "On the quantal nature of speech", J. Phonetics, 17, 3-45.

[11] Bailly, G. and Tran A. (1989), "Compost: a rule-compiler for speech synthesis", *Eurospeech 89* Paris, Vol. 1, pp. 136-139.

[12] Boë, L-J/, (1993), *Speech Maps Interactive Plant "SMIPS"*, Deliverable 6 of SPEECH MAPS, (ESPRIT/BR N° 6975), pp. 3-20.

[13] Stevens, K.N., and Perkell, J., (1993), *Laboratory on the physiology, acoustics and perception of speech*, Course 6542J, 24966J, Massachusetts Institute of Technology.

[14] Maeda, S. (1993), "Acoustics of Vowel Nasalization and articulatory shifts in French Nasal Vowels", in Nasals, nasalization, and the velum, Huffman, M.K. and Krakow, R.A. (eds)., Academic Press, San Diego, pp. 147-166.