INSTRUMENTAL PHONETIC FIELDWORK TECHNIQUES AND RESULTS

Peter Ladefoged

Linguistics Department, UCLA, Los Angeles, U.S.A.

ABSTRACT

Phoneticians can now take much of their laboratory apparatus into the field. Tape recorders have long been available, but their utility is much increased when they are used in conjunction with a portable computer. The computer not only provides convenient editing and play back facilities, but also can produce spectrograms, pitch curves, and other acoustic analyses. In addition, physiological parameters such as pressure and air flow and electroglottograhic data can be recorded and analyzed in the field on a portable computer. Photography (including video recording) and palatography are further tools for field use.

1. INTRODUCTION

There is a story about Daniel Jones, the great British phonetician who dominated the field in the first half of this century. When he was about to go off on a field trip someone asked him what instruments he was going to take with him. He pointed to his ears and said: "Only these." It is surely true that by far the most valuable assets a phonetician can have are a trained set of ears. It is also true (and Daniel Jones would certainly agree) that the ears should be coupled to highly trained vocal organs that are capable of producing a wide range of sounds. There is no substitute for the ability to hear small distinctions in sounds. There is also no substitute for the ability to pronounce alternative possibilities, so that one can ask a speaker which of two pronunciations sounds better. One of the most efficient procedures for getting results in the field is to test different hypotheses by trying out various vocal gestures of one's own. Nevertheless, however well trained they might be, phoneticians who now go out with only their ears and their own vocal apparatus are doing themselves a disservice.

2. RECORDING

What sort of machine should be used for making field recordings? As portable computers become more available, the days of dependence on tape recorders may be passing. Direct recording onto portable computers may be used, with the tape recorder being regarded simply as a backup. The computer system should be capable of sampling speech at 20-24,000 Hz for high quality listening and analysis, and at 10,000 Hz for the analysis of vowels and similar sounds. Even when considered just as devices for reproducing sounds, computers are much more versatile than tape recorders. Fieldworkers want to be able to record word lists or short paragraphs and then to play back selected pieces over and over again, so that they can hear subtle nuances of sounds that are new to them. They also want to be able to hear one sound, and then, immediately afterwards, hear another that may contrast with it. Both tasks can be done somewhat cumbersomely and tediously using tape recorders. But they are trivial, normal operations on any computer equipped with a means for digitizing and editing recorded sounds.

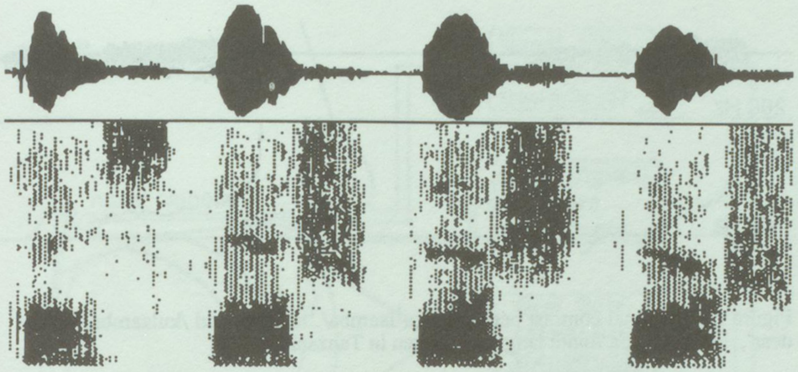

Figure 1. A spectrogram of [kos̪, pos̺, poʃ, poṣ] 'money, milk, language, clan name' in Toda, made under field conditions.

3. ACOUSTIC ANALYSES

In addition to being useful as a sophisticated playback device, a computer can provide several types of analysis that a fieldworker might find useful. The best display of the general acoustic characteristics of a sound is a spectrogram. Figure 1 shows the kind of spectrogram that can be produced on a portable computer without a color (gray scale) screen, printed on a light weight battery operated printer used in the field. The display in Figure 1 was created by a commercially available program, Signalyze. This program should not be judged by the spectrogram in this figure; the spectrograms it can generate on a color screen on a laboratory computer are much more impressive. But even the display that it is possible to print in the field can be very useful. The words shown illustrate the four contrastive sibilants that occur in Toda, a Dravidian language spoken in the Nilgiri Hills in India. Each of these words ends in a different sibilant. The overall spectral characteristics of these sibilants are evident. The laminal dental sibilant at the end of the first word has the highest frequency, and the retroflex sibilant at the end of the last word has the lowest. The apical alveolar and (laminal) palato-alveolar sibilants at the ends of the second and third words have very similar spectral characteristics. (The lowering of the spectral energy peak at the end of the second word is a non-distinctive feature, being simply due to the closure of the lip for the consonant at the beginning of the next word.) These two sibilants are distinguished primarily by their on-glides. The increasing second formant at the end of the third word is due to the raising of the blade and front of the tongue for this laminal sound. In the last word, the lowering of the third formant is probably due to the sublingual cavity that is formed by raising the tip of the tongue for this retroflex sibilant. A great deal of information can be obtained even from these low quality spectrograms, produced under field conditions. Of course, still more information can be obtained from high quality spectrograms produced by this or another program on a laboratory computer at a later date.

Another kind of analysis that is very useful to the fieldworker is one that indicates the pitch. The Signalyze program discussed above will also generate good displays of the fundamental frequency (and it will produce narrow band spectrograms, which are sometimes even more useful for pitch analysis when a creaky voice quality or other unusual spectral characteristics are involved). But a number of other programs will also provide similar information. Figure 2 shows the fundamental frequency in a set of words with contrasting tones in Sukuma, as analyzed by a public domain modification of SoundWave, written at the University of Uppsala, Sweden.

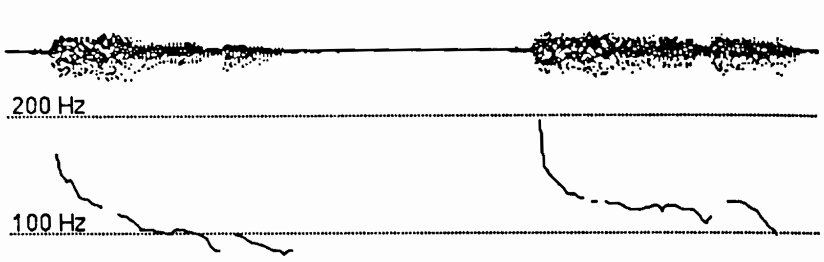

Figure 2. The tonal contrast between /ku!laamba/ 'to lick', and /kulaamba/ 'to be dear', in Sukuma, a Bantu language,spoken in Tanzania.

The final kind of computer analysis of speech sounds that will be discussed here is one for determining the formant frequencies, the principal aspects of vowel quality. A common way of obtaining formant frequencies is by inspection and peak picking using superimposed LPC and FFT displays. The Uppsala software mentioned above provides a convenient way of producing displays of this kind in the field. When making an FFT it is important to remember the system limitations. In effect, an FFT provides the amplitudes of the spectral components that are present on the assumption that these components are all multiples of a wave with frequency depending on the number of points in the FFT. The greater the number of points in the FFT, the longer the wave length, thus the lower the frequency of this wave, and the smaller the interval between calculated components. But any program calculating an FFT will have a certain maximum number of points permissible (usually something like 512 or 256). Accordingly, the only way to further increase the accuracy in the frequency domain (i.e. to decrease the interval between measured components) is to *decrease* the sample rate. This will have the effect of decreasing the range of frequencies that can be observed. But it will also mean that all the components calculated will be within that range. Given a 512 point FFT and a sample rate of 20,000 Hz, there will be 256 components spaced about 40 Hz apart in the range up to 10,000 Hz. But if the sample rate is reduced to 10,000 Hz, the components in the same FFT will be spaced about 20 Hz apart in the range up to 5,000 Hz. It was for this reason that it was suggested earlier that if vowel formants were being studied it is advisable to use a lower sampling rate. The alternative would be to use an FFT with a larger number of points, but no analysis system will permit the maximum number of points to be increased beyond some fixed limit.

4. PHYSIOLOGICAL DATA

Acoustic analyses made from good quality tape recordings can provide large amounts of data. But they often do not indicate in an unambiguous way important articulatory facts such as the degree of nasalization, the phonation type, the direction of the airstream or the timing of movements of the vocal organs. The best way of gaining information on these phonetic parameters is by recording a number of aerodynamic parameters, using a portable computer. The general form of the system we use is shown in Figure 3. We can record the audio signal and up to three physiological signals. Typically these include one pressure (either the pressure of the air in the pharynx obtained by passing a tube through the nose, or the pressure of the air in the mouth using a more convenient tube between the lips), and the oral and nasal air flow. This system provides good data on degrees of nasalization. We have also used it to record an approximation to the subglottal pressure by means of a tube with a small balloon on the end of it in the esophagus, in investigations of prosodic features. Electroglottograhic data can be recorded in a similar way

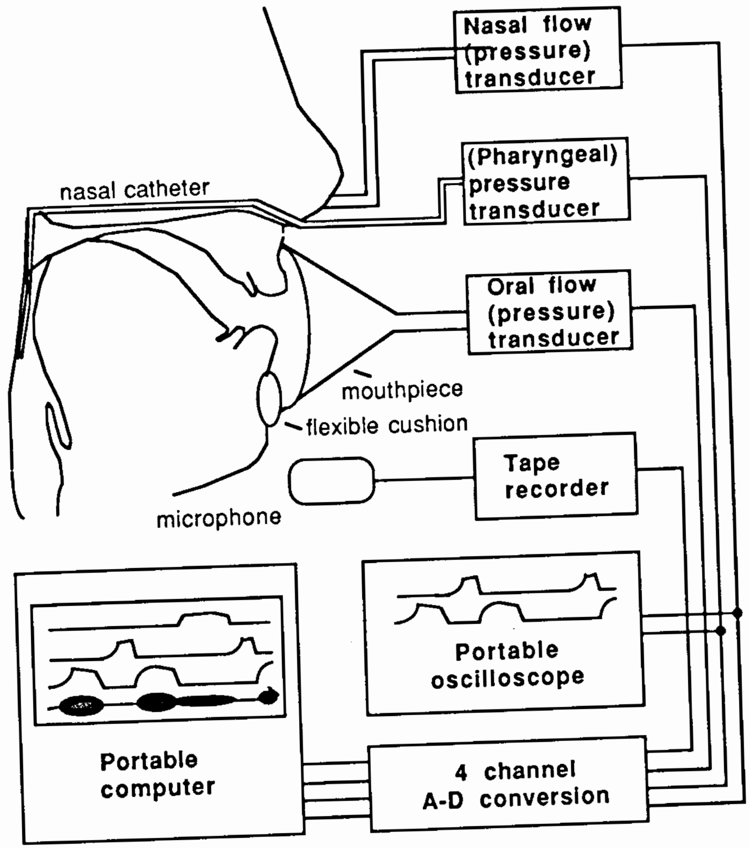

Figure 3. Apparatus for obtaining aerodynamic records in the field.

Fieldworkers want to know not only the manner but also the place of articulation. Photographs of the lips can be very informative particularly if a mirror is used so that a full face and side views are recorded simultaneously. Palatography is also a well known traditional method of obtaining articulatory data that can be used in the field. The comparative simplicity of this technique should not disguise the fact that it is still one of the most useful ways of obtaining information. on the place of articulation and on distinctions between apical and laminal gestures A useful way of recording the (static) palatographic records is by means of a video camera, which can also be used for recording the (dynamic) movements of the lips as mentioned above. Video images can easily be transferred to a computer, where they can be analyzed and measured — all while still in the field.

Finally, it should always be remembered that Daniel Jones was right. All the paraphernalia of the modern phonetics laboratory can never replace the human observer.

My thanks are due to Tony Traill for his wonderful collaboration in an earlier version of this paper.