Chasing ideas in phonetics

Peter Ladefoged*

ABSTRACT

Starting as a poet, I learned about the sounds of words with David Abercrombie. Then, remembering my background in physics, I moved to studying acoustic phonetics and speech synthesis. From there I learned about psychology and how to test perceptual theories. A meeting with a physiologist led to work on the use of the respiratory muscles in speech. Later I landed in Africa teaching English phonetics and learning about African languages. When I went to UCLA to set up a lab I was able to find bright students who helped make computer models of the vocal tract and taught me linguistic theory. And I was able to continue wandering around the world, describing the sounds of a wide range of languages.

I was nearly 14 in 1939 when World War II broke out. By Christmas of that year my 20 year old brother's ship had been reported missing with all hands lost, and three years later, in 1942, my other brother's plane had fallen from the sky. But I was still at school, writing poems while studying physics, chemistry and mathematics. As the government thought it needed scientists, I was sent up to Cambridge instead of being allowed to go into the army.

At Cambridge I did all the appropriate things, like falling in love with a beautiful woman, and writing terrible poems for her. She wanted to marry a clergyman, so I converted from nothing to clerical Christianity and found a bishop who took me on as an ordinand of the Church of England. Unfortunately I was so busy with all these activities that I didn't learn much science. At the end of the year I wrote my biochemistry exam in rather bad verse, which is perhaps why the examiners sent me off into the army.

When I came out of the army I wanted to find out what made great poetry, so I went to Edinburgh University. Within my first year I found that nobody in the English Department was teaching anything about poems and the sounds and rhythms of words. They were all more concerned with what the poet meant, rather than why

* Phonetics Laboratory, Linguistics Department, UCLA

one set of words sounded better than another. Fortunately I chanced on someone in the basement, David Abercrombie, who had just started a new thing called a Phonetics Department. He was himself the son of a poet, Lascelles Abercrombie, and he taught me how to listen to what people said, and how speech sounds were made.

After two years at Edinburgh I was entitled to a first degree, not because I was smart but because ex-servicemen were allowed a year off from the usual three year requirement for an Ordinary degree. And because the Scots are a canny people who don't want to waste time on frivolities, they don't give BA degrees. I got an M.A. (ordinary) (war emergency) in 1951, and was able to go on to do postgraduate work in phonetics. A year later I had my first job, as a lab assistant cutting vinyl recordings; and on January 1st 1953 I was promoted to Assistant Lecturer in Phonetics.

In 1952 I noticed that the VIIth International Congress of Linguists was being held in London, and asked David Abercrombie if I should attend. He said that, personally, he wouldn't cross the road to go to such an event, but if I were visiting my parents who lived near London, I might as well look in. I am glad I did, because in an obscure session that a dozen people attended Walter Lawrence demonstrated the world's first parametric speech synthesizer. Frank Cooper had built the pattern playback at Haskins, which reproduced sound spectrograms, but Walter Lawrence was the first to combine a varying voice source, three formant resonators and fricative noises, to produce synthetic speech. I was intrigued and had a long conversation with him afterwards. It turned out that he was looking for a university that would undertake research using his synthesizer. He came up to Edinburgh and was charmed by David Abercrombie, delighted with the promise of good technical support from Jim Anthony, and pleased that I wanted to do work on speech perception.

Walter Lawrence suggested that we work with a Cambridge psychologist, Donald Broadbent, who was also interested in the use of synthesized speech in perceptual experiments. Together we did an experiment in which we synthesized a voice saying 'Please say what this word is:' followed by a synthesized version of one of the words 'bit, bet, bat, but.' We were able to show that the preliminary sentence had a noticeable effect on the perception of the following word. If all the vowels in the introductory sentence had a comparatively high first formant, then the vowel in a stimulus word will appear to have a lower first formant than it does. A word with a vowel halfway between 'bit' and 'bet' can sound more like 'bit' on one occasion, and more like 'bet' on another. depending on the introductory sentence.

We now know that varying the context in which a word appears is much less important in real speech than in our experiments. When listening to poor quality synthesized speech, subjects use all the information they can get including that in the

surrounding words in order to decide what was said. But when listening to high quality natural speech there is more information in the test word itself, and listeners pay less regard to the surrounding context. It was a fun experiment, but less illuminating than we thought at the time.

Donald Broadbent and I worked together on a number of other perceptual problems that may have more far reaching implications. I had noticed that when playing a recording with a click superimposed on an utterance, it was often difficult to tell where the superimposed click occurred. In an experiment in which we superimposed an s - a regular speech sound - on a recording we found that listeners were most accurate in their judgments when the s occurred on a stressed syllable. When it was elsewhere they usually thought the s had occurred earlier, sometimes by as much as one or two syllables. We also found that naive listeners were not very skilled in perceiving the correct sequence of auditory events. especially when they could not produce the sounds themselves. Anyone can tell whether an s comes before or after the vowel in a word such as 'say' or 'ace'. But when considering a sound made up of a whistle and a buzz of comparable lengths to the s and the vowel in 'say', naive listeners usually cannot tell whether the whistle comes before or after the buzz. Both Donald Broadbent and I held, as I still hold (Donald, alas, is dead), that when listening to normal speech people decode chunks the size of a syllable or greater. Speech perception does not involve a segmental matching process.

Speech perception was not my only concern at that time. In the mid 1950's, while I was working on these perceptual problems, David Abercrombie, asked me to dinner with David Whitteridge, a well known physiologist. Abercrombie told him that we were interested in Stetson's theory that every syllable was accompanied by a 'chest pulse' produced by the internal intercostal muscles between the ribs. Stetson also thought that stressed syllables had a 'reinforced chest pulse' in which the rectus abdominis muscle in the abdomen reinforced the internal intercostal action. Another point of Stetson's theory was that open syllables (those without consonants at the end) were checked by the action of the external intercostals, a second set of muscles between the ribs. On hearing all this, David Whitteridge turned to me and said "Why don't you come along to my lab on Saturday morning, and we'll soon check this out." It took us only a few Saturday mornings to realize that this was not a simple task and that we had to undertake a large research project, one that went on for several years. David Whitteridge was a great researcher and it was from him that I learned how to be a scientist, how to keep proper records, how to measure things, the importance of calibration and all the rituals of science.

We made a number of errors. The most far-reaching was that we did not understand about esophageal pressure. We thought that a suitably placed balloon in

the esophagus would record the equivalent of subglottal pressure, which is incorrect. There is a complex, non-linear relationship between the true subglottal pressure and the pressure in a balloon in the esophagus just below the vocal folds. We also had 1950's equipment. We recorded the pressure on an ink writing kymograph by means of a rubber tambour made from a condom. There is the well known and slightly apocryphal story of our getting an assistant to go out and buy several boxes of condoms. He was asked what kind of condoms he wanted, and reputedly answered: "It doesn't matter. We cut the ends off before we use them." (There is, incidentally, no truth to the rumor that the same lab assistant became even more famous because he went out and bought a box of a dozen condoms every day for a week.)

We measured the air in the lungs by using a body plethysmograph made by welding a couple of oil drums together. Speakers were sealed into the barrels by a tightly fitting collar around the neck. As they breathed in and out and spoke our test phrases, the air displaced from the barrels was measured. Getting the speaker placed so that the collar was in the right position was always a problem. His position was adjusted by varying the number of telephone directories he stood on at the bottom of the barrel.

Electrodes were inserted into the appropriate muscles through ports in the barrels which then had to be sealed up. The electrodes were made by taking a pair of varnished copper wires unwound from a solenoid and passing them through a fairly large hypodermic needle. The exposed tips were ground flush with the slanting tip of the needle. One of my first memories of these experiments is seeing David Whitteridge, a somewhat overweight middle aged man, pushing the needle into a muscle close to the belly button, and saying "This needle is rather blunt. I'll have to use it like a drill." He then spun the needle between his fingers while pushing it hard into his ample belly.

Despite all these problems we got some useful results. We were able to determine that Stetson was wrong. Syllables cannot be defined in terms of pulses from the internal intercostals. Stressed syllables do not necessarily have an additional push from the abdominal muscles; and the external intercostals are not used to check all open syllables. We also found that, depending on the amount of air in the lungs, different muscles are used to produce similar sounding utterances, a phenomenon that is now known as motor equivalence. For most of our speakers the external intercostals provided a checking action during the first part of an utterance after an inspiration. The internal intercostals pushed air out of the lungs as the volume decreased; and additional expiratory muscles came in later. This pattern, which we noted in the 1950's, was later confirmed by the British physiologist Tom Sears. His work also shows emg records of the inspiratory muscles checking the descent of the rib cage at

the beginning of phonation, and then the expiratory muscles coming in with increasing activity as a long sustained tone is produced.

Recently Tom Hixon and Gary Weismer have pointed out how completely wrong we were in our interpretation of our pressure records. They have also shown that we underestimated the use of rectus abdominis for many speakers, and even wrongly characterized its use in our own records. However, the modified report in my *Three areas of Experimental phonetics*, which sums up most of my Edinburgh work, still correctly describes the use of the respiratory muscles in the speech of many people. The overall picture of the use of different muscles at different lung volumes remains valid, as long as we remember that there is a great deal of variation in the different parts of a muscle as used by a single speaker, and there is considerable variation between speakers.

In the late 1950's I thought it might be nice to work in the United States. This required my completing a Ph.D., a degree that most of my colleagues in the Faculty of Arts did not have. I went to the University Registrar and asked him what I should do. He noted that I'd been a member of the faculty for more than three years, which would count as the period that I needed to be registered. All I needed was a thesis. So I consulted David Abercrombie, and on his advice took three papers that I had already published on aspects of vowel quality, and added an introductory survey. I also appended some work that I had been doing on Cardinal Vowels with Daniel Jones, who had recently retired from the chair of phonetics at University College, London.

Abercrombie had arranged for a grant so that Daniel Jones could be a consultant on a project to study the acoustic quality of cardinal vowels, largely so that I would have the opportunity to work with the leading phonetician of the time. The assumption was that cardinal vowels would always have the same quality even when produced by different people. With hindsight I can now see that this was not the case, as there are ambiguities in the definitions of cardinal vowels. One part of the definition says that they should be auditorily equidistant vowels, another that they should be imitations of the vowels of a teacher who knows them, with Daniel Jones being assumed to be the original teacher. Most observers consider the Daniel Jones vowels to be auditorily closer together in the back series than in the front. The sets of vowels that I recorded were all produced under Daniel Jones's immediate supervision by his former pupils. They were probably imitating him to their best ability; but they may also have been influenced by their own notions of equidistant vowels. The result of this part of the thesis are not particularly noteworthy, except insofar as they provide an early example of the problems of analyzing vowels spectrographically.

Of course, when I had completed the thesis I had to have an oral exam. This had

a slightly unusual form. The outside examiner was Walter Lawrence, who came up from London and said "Well, Peter, I haven't actually read the whole thesis yet, but I know the published papers. Let's go off to a pub and chat for a bit. Then I'll sign the forms." And that was how I added a Ph.D. to my M.A. (ordinary) (war emergency).

Before going to America, I returned to Nigeria. I had already been there for a year, on a leave of absence from Edinburgh, attached to the new Phonetics Department at the University of Ibadan. I was amazed by the large number of different languages, and the opportunities for fieldwork. I had had some fieldwork experience in Scotland. Soon after my M.A. David Abercrombie had sent me off to the Gaelic speaking Outer Hebrides. He told me to find out about the differences between pairs of words such as *dubhan* **tu.an** (hook) and *duan* **tuan** (song). Many years later I went back to the Hebrides, and noted the pitch differences between these words, the first of which counts as two syllables, and the second as one. But this first field trip was an utter disaster. I had no idea how to elicit phonetic data, and came back after only a few days, having been able to find out nothing. A year or so later I had a much more profitable experience working on the Linguistic Survey of Scotland. Ian Catford had devised a well structured task in which fieldworkers had to elicit a set of words designed to make the phonology of different dialects apparent. I still had the problem that I was not good at walking up to strange houses and asking the inhabitants if I could study their vowels. But fortunately by then I had married Jenny, a much more talented and wonderful woman than any I had ever known before, and she had no difficulty in making the initial contacts with suitable speakers. We got some good data on three Scottish dialects, and I learned the first steps about eliciting phonetic data.

When we returned to Nigeria I was a field fellow on the Linguistic Survey of West Africa, a Ford Foundation sponsored project, led by Joe Greenberg and Bill Welmers. Using the talents of university students in Nigeria, Ghana, Ivory Coast, Sierra Leone and Senegal I was able to record speakers of about 60 different languages, and to make palatographic and aerodynamic studies of many of them. The principal results appeared in *A phonetic study of West African languages* (1964), a book that I now recognize as flawed because it so often described a language based on the analysis of only a single speaker.

In 1962 I was appointed as an Assistant Professor of Phonetics at UCLA, and less than a year later received an NIH grant to build a working model of the vocal organs. This was the start of the UCLA Phonetics Lab. Jim Anthony came over on leave from the University of Edinburgh, and we worked with rubber molds and plaster casts, and failed to achieve anything noteworthy. We were saved from ignominy by the contributions of Vicki Fromkin, John Ohala and other colleagues who were using

electromyography and aerodynamic techniques to describe the muscular activity and associated gestures that we had hoped to build into the model.

In 1967 the UCLA Phonetics Lab acquired a LINC-8 computer, described in the lab report as "a large general purpose computer with 8K of memory." Richard Harshman, who did all the early programming, performed wonders with this minuscule memory (including teaching me how to program). Later the computer was upgraded to a machine thought of as having a massive memory, totaling 32K, enabling Lloyd Rice to help us salvage the failure to build a physical vocal tract by programming a computer model.

Our articulatory modeling was based on x-ray views of the vocal tract. This gave us the problem of converting two dimensional sagittal views into three dimensional data incorporating the width of the vocal tract as seen in the other dimension. We tried to make a physical model of the vocal tract in the position for a neutral vowel, using dental impression material. This is not difficult to do for the oral cavity. But we needed in addition an impression of the pharynx. With the help of a cooperative dental surgeon we achieved this by filling my vocal tract with a quick setting impression material. I was turned upside down, so that the impression material could be inserted *up to* the tops of my arytenoid cartilages. I only had to hold my breath for 30 seconds. We made a nice model of this actual vocal tract shape but unfortunately it was not very accurate as the weight of the impression material distorted the pharynx, making me appear to have too large a pharyngeal cavity.

We also had to specify the vocal tract shapes for a variety of sounds. We needed a way of characterizing possible tongue shapes. We obtained x-ray data on the tongue positions in the middle of each of the English vowels as spoken by 5 speakers. While working in the UCLA Phonetics Lab Richard Harshman had invented PARAFAC, a form of factor analysis that provided a unique set of factors underlying the variation in such data. Using PARAFAC we were able to show that two factors, front raising and back raising, could generate most of the possible gestures of the tongue body. A number of computer models now use these two factors (or variants of them) to specify tongue body shapes. We also found that we could use these factors for deriving vocal tract shapes from formant frequencies. Studies of the different articulatory gestures used by different speakers led us to conclude that, alongside the possibility of a motor theory of speech perception, there is support for a theory showing that at least part of the speech output is controlled in auditory terms. In other words, we have an auditory theory of speech production.

UCLA allowed me to create my own phonetics courses. They were similar to those taught at Edinburgh, but with far more emphasis on learning to distinguish the sounds of other languages. The capstone of the introductory course was the requirement to

make a recording and write a paper describing the sounds of another language. Students found this a challenging task; and I was able to work with them on numerous languages, using their friends, aunts, uncles and room mates to provide an extraordinary range of data, the best of which joined my own field recordings in the UCLA phonetic archive.

I was also able to create a Phonetics Laboratory group. For me, the people mattered more than the instruments. Of course we tried to get good computers and all the hardware we needed. But getting a group that functioned together and saw themselves as a working unit was of prime importance. I enjoyed going around every day, chatting to everyone, and lunching in the lab with as many people as possible. It was also useful to have weekly lab meetings which all the staff, students and faculty were expected to attend, even if only for the first ten minutes when we talked about what was going on. Building a research group who felt that they had a stake in the development of the lab taught me their varied ideas from statistics to engineering, and the philosophy of linguistics. And interacting with bright students keeps one intellectually honest.

Throughout my years at UCLA I have spent much of my time wandering around the world trying to hear and analyze all the sounds that could distinguish words in some language or other. To begin with I had a portable phonetics lab which required a porter. It weighed more than 100 pounds, and included a Nagra tape recorder, a battery powered oscilloscope, and an ultra violet recorder, plus all the paraphernalia required for palatography and pressure and flow recording. Nowadays I can get by with much less. The heart of the apparatus is a small DAT recorder, and a solar powered computer. Air pressure and flow measurements are made with battery operated equipment. We have multichannel software that provides spectrograms, LPC and FFT spectra, and pitch and intensity displays so that I can make complex analyses in the field. I now use a video camera not only for recording dynamic movements of the lips but also for static palatography.

I have enjoyed wandering to many corners of the earth, though fieldwork has not always been comfortable. I remember once sitting in a small boat in the Niger delta, made for perhaps 12 people. The 24 of us crammed in there were huddled under a ground sheet as torrential rain was pouring down. I had my expensive tape recorder and microphones in a theoretically waterproof bag in the bottom of the boat, with the water slowly rising. Wet and worried I wondered whether our insurance really covered the thousands of dollars of equipment. But later we sat in the village chief's hut, poured a libation of some strange potent liquor, and recorded a dozen speakers of Defaka, a dying language spoken by only a few hundred people on one of the islands in the Niger delta. When the skies had cleared we went back in an old dugout canoe. Warm and

dry I watched the sun setting, thinking how lucky I was to have these opportunities.

Another delight of fieldwork is the charm of the people one meets. The Xóõ, who were willing to have tubes put through their noses; the Hadza who have fewer possessions than anyone I know, except perhaps the Pirahã, who live with little thought for the morrow; the Toda whose courtesy and helpfulness were unparalleled; the Tsou, who could not understand why anyone would come to their mountain to record their sounds; and all manner of peoples from the Aleutian Islands to the Australian outback.

Very little of my fieldwork would have been possible without the cooperation of many great linguists. Most notable among them are Tony Traill and Jan Snyman, who took us to parts of Namibia and Botswana that we would never have visited on our own; Bhaskararao, who guided us through India, looking after our food and drink, as well as our linguistic needs; Kay Williamson, who demonstrated many of the phonetic treasures of Nigeria; Dan and Keren Everett who led me through the Amazonian rain forest; and scores of other local linguists who have been helpful, including many members of the Summer Institute of Linguistics and other missionaries who did not worry about working with a member of Atheists for Jesus.

Fieldwork allowed me to make quantified observations of a wide variety of sounds. David Whitteridge was fond of quoting Lord Kelvin to me: "You do not really know anything until you can express it in terms of numbers." Or, as Jenny Ladefoged puts it, equally succinctly: "Numbers are a scientist's security blanket." Sounds that I have been able to investigate include several not previously measured and numerically described in the phonetic literature, such as voiceless implosives, bilabial trills, and velar laterals. Some of my nicely quantified results have later been found to be inadequate, often because I had not realized the necessity of studying several speakers of a language. Two people saying a set of phrases six times are nothing like as useful as six people saying the same set of phrases twice. There is far more variation between speakers saying the same thing than between repetitions of the same utterance by one speaker. But despite the limitations of some of the work, many of the instrumental observations have been found to be useful.

Phonetic fieldwork and the study of the sounds of a wide variety of languages led to the construction of linguistic phonetic feature systems. It was not difficult to see that the Jakobsonian and later SPE feature systems were inadequate. In *Preliminaries to Linguistic Analysis* (1971) I proposed a new feature system that accounted for a wide range of phonation types and several articulations previously unknown in the phonetic literature. This work also demonstrated the importance of recognizing an auditory basis for some features. Chomsky and Halle thought that features could be defined equally well in articulatory or auditory terms. This is incorrect. Some features

group sounds together because of their auditory similarity and others group sounds that have similar articulations.

Fieldwork studies were part of my impetus for urging revision of the IPA. The IPA chart is like a one page theory accounting for all the possible speech sounds that can contrast in a language. By the mid 1980's it was apparent that the existing chart was out of date, and the International Phonetic Association needed to catch up with the times. As president, I was able to help by convening the 1989 Kiel Convention, which led to major revisions of the IPA chart.

Much of what I have learned about the *Sounds of the world's languages* is included in the book of that name, which I wrote with my colleague Ian Maddieson. Another of the good chances of my life is that I have been able to work with him. He, too, has first-hand knowledge of the speech of many lands. But his talents are different from mine. He has a more thorough, scholarly approach. He was able to curb my wilder flights of fancy, so that we produced a more balanced book, summing up the present state of our knowledge of the sounds of the world's languages.

Nevertheless. I'm still chasing ideas in poems and phonetics.

I grow old but do not worry
That I've passed the painful time
When pride makes us hurry
To find the needed rhyme.

Age creaks my bones. I know
No hope that I can reach
A land where rhythms grow
Beside a lyric beach.

But I'm happy still to wander
Along the sound-filled coast,
Twisting words to ponder
Which it is that pleases most.
Hoping sweet sounds will not be gone,
And wisps of songs will linger on.

Received : March 22, 1999

▲ Phonetics Laboratory
Dept. of Linguistics
3125 Campbell Hall
UCLA, Los Angeles,
CA 90095-1543
e-mail: oldfogey@ucla.edu