

PHONETICS, PHONOLOGY, AND THE TEACHING OF PRONUNCIATION — A NEW CD-ROM FOR ESL LEARNERS, AND ITS RATIONALE

Helen Fraser
University of New England, Armidale, NSW, Australia

ABSTRACT

As phoneticians, it is relatively easy for us to describe in articulatory or acoustic terms the aspects of a learner's speech which are different from those of a native speaker. It is also becoming more feasible to automate judgement of a learner's utterances in comparison with those of a native speaker, providing instant feedback to the speaker via a computer screen.

Questions remain however. Which of the many differences between a native and non-native pronunciation to bring to the attention of the learner? Even more importantly, how can those differences be presented to learners so that they can actually make use of the information to change their pronunciation appropriately? For example, though it is obvious that instructing a learner to 'raise the second formant higher when you are 80% through the second vowel' will be of little use, it is far less obvious what metalinguistic descriptions *are* useful to learners.

In this paper, I put forward some principles developed through both theoretical and practical investigation to help ensure effective metalinguistic communication between teachers and learners of pronunciation. I demonstrate a CD-ROM based on these principles which I have recently produced to test experimentally a range of methods of helping ESL learners with pronunciation.

INTRODUCTION

Over recent years there has been an explosion in research and development aimed at creation of automatic pronunciation tutors for learners of English as a second language. The goal is a computer-based product which can analyse learners' pronunciation of English phrases, and give feedback, usually in terms of a score, on how closely the learner matches a native speaker model.

This seems intuitively to be a good idea, allowing learners to practise pronunciation even in the absence of a human tutor. However, as many linguists and others are aware, it is not quite as simple as sometimes thought. Apart from the technological problems involved in making such computer applications work, there are difficult linguistic issues to address: Which aspects of a learner's speech to focus on? How much deviation from the model to tolerate? After all, most learners, even successful ones, will speak with a 'foreign accent' somewhat different from that of native speakers. It is important to concentrate on those aspects of learners' speech which affect their intelligibility, not to pick them up on every slight difference from a native model.

However, to the extent that they can be made to mimic the judgement of a human tutor — and some of the better ones show promise in achieving this — such automatic tutors may be useful, and certainly have a place in a suite of language learning tools.

It is important nevertheless to deflate the high hopes that are sometimes invested in such tools, by raising some less obvious issues. Even when automatic tutors do work effectively, there remain learner needs that they cannot satisfy.

Firstly, learners not only need to know that their pronunciation was incorrect; they also, crucially, need guidance on how to improve.

Secondly, learners must ultimately gain the ability to make their own judgments about English pronunciation, their own as well as others', rather than relying on external assessment. In other words they need, as part of the language-learning process, to come to hear English as native speakers do.

In my own work, I use technology for teaching pronunciation in a way which addresses these issues. This is the topic of the current paper.

RELATIONSHIPS AMONG LEARNER, TEACHER, AND COMPUTER

The broader issue perhaps recapitulates in a modified way the old debate about Artificial Intelligence (AI): should AI aim to create a machine that does what human intelligence does, or should it rather aim to create a machine that interacts with human intelligence in order to better achieve the goals of humans (cf. Winograd and Flores 1986)? In the present context, the question is, should we aim to create a computer application that mimics the role of the teacher, or should we rather use technology to enhance the interaction between learner and teacher? In my work, I take the latter approach, attempting to use the many opportunities afforded by recent technological developments to extend the ways in which a teacher can interact with learners, both in the classroom and at a distance.

HEARING VERSUS SEEING

The key opportunity afforded learners by computer technology, in my view, is the ability to hear both their own pronunciation and a model 'externally' and repeatedly. This allows learners to study for themselves the differences between their pronunciation and the model, in a way that is next to impossible with real-time speech, and gives them the responsibility of learning to judge the differences between the two versions, rather than relying on external assessment.

Note that this view is at variance with the common approach of providing learners with a visual representation of their utterance, in either articulatory or acoustic form. This can be done in a wide range of ways, some far more effective than others. Even the best, however, rely on the assumption that the learner directly controls, and is able to modify at will, the articulation and acoustics of their speech.

This may seem a self-evidently valid assumption, but I believe it is in fact invalid. Speakers are not aware of, and consciously controlling, the articulation and acoustics of their speech *as such*. Rather they are aware of and consciously controlling the auditory output of their speech. Even though learners can observe, and some can even understand, articulatory and acoustic representations of speech, they do not have a deep intuitive understanding of how to use these to control their speech. Articulatory and acoustic descriptions of what is going on are secondary products, which only make sense on the basis of relevant background knowledge of linguistics and phonetics, rather than in an intuitive, 'knowing how' way.

The situation is quite different for auditory perception of speech. To a large extent people do have a deep intuitive understanding of how to manipulate their vocal organs to create a particular phonetic output. I suggest that when articulatory and acoustic representations of speech in automatic tutors (appear to) 'work', this is in cases where learners actually hear the sound as well as seeing the representation. In cases where there is no auditory version, but only, say, a pitch trace or mid-sagittal section, the learner is left completely in the dark [demonstration in the oral presentation].

The point here is, knowing a lot about articulatory and acoustic phonetics is something that helps linguists, but doesn't really help learners. What 'drives' pronunciation is not an articulatory or acoustic representation, but an auditory image'.

WHAT TYPE OF FEEDBACK TO GIVE LEARNERS?

Of course, simply listening to oneself and a model is not in itself sufficient to ensure improvement. Two other things are needed: guidance for the learner in interpreting the *significant* differences between the model and the attempt; and understanding by the learner of what to do in order to come closer to the model on those significant aspects.

A large part of the reason learners have problems with English pronunciation is that they have not yet conceptualised the distinctions that are significant in English, as opposed to their native language.

Only a minority of learner difficulties arise from actual inability to produce a particular sound, and those difficulties are generally the least significant from the native listener's point of view (eg. inability to pronounce a good dental fricative interferes relatively little with intelligibility). Much more significant are difficulties in noticing and controlling the use of sounds and syllables that are not in themselves a problem for the learner to produce (eg. many learners who find it quite easy to say 'toff' find it quite difficult to say 'put off'). Learners, then, need a little help with perceiving and producing of English sounds, but a lot of help with conceptualising sounds in an English way (see Wierzbicka (1990) for a nice discussion of this distinction in relation to the description of colours, an area whose issues are remarkably similar to those of phonological representation).

PHONETICS, PHONOLOGY, AND PRONUNCIATION

However, giving learners an auditory model, as we have already seen, is not in itself enough to help them improve their pronunciation. For a short stretch of speech, it is possible (at least for some people; others find it much harder) to produce a purely 'phonetic' repetition. However in order to be able to *use* the pronunciation — remember it, reproduce it, alter it appropriately — learners also need some way of 'thinking about' the auditory image; in other words they need some kind of linguistic or phonological conceptualisation of the sounds they are hearing.

This raises questions which do not have clear answers in psycholinguistics and speech production theory, and in fact are not often addressed in these fields. There are a wide range of different approaches to phonological and phonetic representation in speech production, differing with respect to issues like the size and nature of units of representation at different levels of processing, the relevance of various types of processes relating them, and so on (see Lófqvist (1997) for a brief review and references). Almost all of these however address issues of the *subconscious* control of speech by speakers who 'know what they want to say' — ie. know the auditory output they need to produce in order to send their message. Far fewer theorists address the question of the *conscious* control of speech — how to decide upon the right auditory image in the first place, prior to implementing whatever type of subconscious plan it is that people use to realise their speech intentions.

Consider an example. A classic type of speech production experiment discussed by Lófqvist (1997) uses the phrase 'It's a papaya', and considers the complex subconscious control mechanisms involved in moving all the articulators in the right order and manner to create this output. The issue this does not address (and does not pretend to) is 'what if the experiment were done with a non-native speaker, who put the stress on the wrong syllable, produced one /p/ so that it sounded more like an /f/, and got the vowel quality wrong?'. Of course it would be possible to describe the differences of timing and gesture that contribute to these differences of output. However to think that such a description accounts for the whole of the learner's problem would be to miss something very important (though it might be useful in understanding some types of 'low-level' speech pathology).

Thus the area of second language pronunciation makes a particularly challenging testing ground for theories of speech production, since we need actually to be able to help learners modify their pronunciation, not simply theorise about what goes on under the surface of fluent speech.

I believe that the most serious problem second language learners face is in the forming of an appropriate conscious intention; the subconscious realisation of their intention is also a problem, but a minor one by comparison. Once the learner can form the right intention, good pronunciation is merely a matter of practising a skill; without the right intention, no amount of practice will result in good pronunciation.

Speech production theory has had very little to say, explicitly, about the conscious formation of the intention to say something. However, implicit assumptions about such intentions underpin much theorising. The problem is that some of these assumptions, when made explicit and examined are found to be quite erroneous, and don't work at all well when challenged to help with the practical problems of second language pronunciation. So, though it is tempting to import ideas from speech production theory into the study of second language phonology — and of course much knowledge from these areas is highly relevant — it is important to note the limitations in its application.

Firstly, the learner's problem is not a general inability to control oral gestures to create an intended auditory image. Note that this is not to deny the importance of practice and skill development in the learning of pronunciation — indeed 'pronunciation is a skill similar to sport or music' is a strong theme of all my work on pronunciation teaching. It is to emphasise that teaching learners 'how to make the sounds' is only a small part of teaching pronunciation.

Secondly, the learner has no insight into low-level, subconscious speech control mechanisms, and cannot modify them at will. In order to help the learner modify their pronunciation, we need to use much higher level descriptions of the differences between their production and the target production. This brings up the topic of phonological representation.

PHONOLOGICAL REPRESENTATION

Again, issues of phonological representation are complex and theory-laden. Again, the issues involved in helping learners are ones not generally addressed in the vast literature on phonological representation. And again, work on second language phonology involves major challenges to implicitly held, but erroneous, underlying assumptions of much mainstream phonological theory.

In particular, it raises very starkly the question of 'whose phonology?' — should the speech be represented in terms of the phonology of the learner's native language? The phonology of English? Some kind of mixture of these? If so, what kind? Some kind of 'universal phonology'? Are the same answers valid for the representation of the learner's attempts as of the model? For the model as perceived by the learner? For the model as understood by the native speaker? These kinds of question really require detailed consideration of the nature and meaning of 'representation' as a general concept — an issue which is very rarely addressed in the cognitive disciplines which use the concept as a corner stone of theory and practice. Once these issues of choosing a phonological representation relative to a particular class of language user (learner, native speaker, etc) are raised, many other questions need to be asked, and some rethinking of fundamental assumptions becomes imperative (see Fraser 1992, 1997b).

This is not the place to go into any detail on these issues. However it is worth mentioning one practical application of these theoretical ideas in an area quite analogous to the present one of second language pronunciation — that of dictionary pronunciation guides. Adherence to the 'obvious' presuppositions of mainstream phonology gave rise to a virtually unquestioned belief that dictionary pronunciation guides should be most effective if given in terms of a systematic, unambiguous code such as a phonemic transcription. As demonstrated and discussed in Fraser (1997a) this is far from being the truth, and such guides are so ineffective that they are rarely even consulted by ordinary dictionary users.

RELEVANCE FOR TEACHING PRONUNCIATION

Through consideration of these issues, I have developed a framework for understanding pronunciation and phonology which, interestingly, fits well with a communicative approach to language teaching (widely held to be the best methodology), but extends that approach into the area of pronunciation — a topic often neglected or inadequately taught in ESL classes (Fraser 2000a). Making this extension requires consideration of three distinct ways in which pronunciation lessons need to be communicative to be effective (cf. Fraser 2000d):

- a) lessons should focus on 'real speech in real situations';
- b) lessons should encourage the learner to think of pronunciation as communication, and consider explicitly the needs of the listener, who uses clues in the learner's speech to understand the message;
- c) lessons should emphasise the need for good metalinguistic communication between teacher and learner about pronunciation. In particular, care should be taken not to frame explanations or suggestions in terms which presuppose intuitive understanding of English phonology, which learners

by definition do not have. In order to help learners, we must understand how they presently conceptualise sounds, and give them tools to help them modify their present concepts in a direction more appropriate to English phonology.

DESCRIPTION AND DEMONSTRATION OF THE CD

The CD-ROM *Learn to Speak Clearly in English* (Fraser 2000b) attempts to embody some of these ideas in a language learning technology, so as to test their efficacy for practical pronunciation teaching.

Its communicative approach is evidenced by the first section on communication, which helps learners think about their pronunciation in terms of what their listener is getting from it, rather than what they are putting into it, and explains in simple terms what it is they need to do to learn the pronunciation of a new language, and why it seems so difficult to them. The order of the next modules follows the order of importance to effective communication in English — first, sentence stress, then word stress, then segmental production.

The approach throughout is 'Don't tell them, show them'. This allows them as much as possible to form their own concepts of English sounds, rather than having alien, confusing concepts thrust upon them.

First they are trained to hear English stress through a combination of visual and auditory feedback, using a normal orthographic representation — ie. a sentence or word is played, and simultaneously displayed, with its stressed syllables visually highlighted. Then they are encouraged to practise their own production, through recording their own attempt and comparing it auditorily to a native speaker model — clicking to play each as often as they wish.

The last module is an entirely new type of exercise, called 'Critical Listening' in which users listen not to models of native speakers, but to other learners. They are asked to critique the learner's pronunciation, in terms of whether it would be easy or difficult for an English listener to understand. Once they have tried this themselves, they can click to hear a native speaker saying the same thing, click to get some notes on the difference between the learner and native speaker, and click to record their own version to compare with the two pre-recordings.

Note that the feedback provided uses descriptions of sounds which we can be confident actually mean something to learners. For example, the concept of 'stress' has just been thoroughly taught — much more thoroughly than is usually done. This is very important, since many learners do not come equipped with an intuitive understanding of this concept, and therefore are apt to miss the point of many lessons unless it is first made very clear to them. Particular sounds are referred to mainly by the letter which represents them in the spelling of the particular word in question. Demonstration of the effectiveness of this practice needs further research, but it seems informally to be very useful to learners, parallel to the work already mentioned on dictionary pronunciation guides for native speakers of English.

CONCLUSION AND NEXT STEPS

The technology described here aims to mediate between the teacher and the learner, and to create demonstrations that are difficult to arrange in the average face to face class. In no way does it aim to replace or mimic the teacher's input. Neither does it aim to impart the teacher's (or phonetician's) knowledge directly to the user. Rather it aims to embody knowledge of linguists, phoneticians and teachers and embed this knowledge in a product that allows learners to do what they need to do in order to learn to pronounce English effectively: reconceptualise speech in terms appropriate to English, and practise the articulatory skills required to transform their new concepts into sound fluently and confidently. This emphasises the need, not always as obvious as it should be, for pronunciation tutoring technology to be produced by teams with real expertise in linguistics and language teaching, as well as technical wizardry.

A second aim of the CD is to demonstrate to teachers some methods of teaching pronunciation that are seldom used, with or without computer technology, since unfortunately most language teachers (with notable exceptions) have little training or confidence in teaching pronunciation, and those that do often take a non-communicative approach, eg. one based on imparting rules of English phonology. But that is another story, currently being addressed in a new CD intended for teachers themselves (Fraser, in prep).

NOTES

Note that this is not the blatant contradiction of Motor Theory or Event Perception it may seem: these models suggest that we interpret an auditory image *as or in terms of* its articulatory production; they do not suggest that we do not require the auditory image at all. The present argument emphasises only the necessity of the learner receiving an auditory model, and is not speaking, at this stage, to the issue motor theorists and others debate: whether that auditory model is interpreted in terms of articulatory gestures, acoustic targets, or something else. There is not space here to go into more detail about the relationship between these models and the views put forward here (though see bibliographic references) — however, as I argue in the next section, there are good reasons to consider that learners use a linguistic conceptualisation, rather than any of those just mentioned.

REFERENCES

- Fraser, H. (1992) *The Subject of Speech Perception: An analysis of the philosophical foundations of the information-processing model of cognition*, (Macmillan: London)
- Fraser, H. (1996) "An Introduction to Phenomenological Phonology (PP)", P. McCormack and A. Russell. Proceedings: Australian Speech Science and Technology Association meeting, Adelaide. (Australian Speech Science and Technology Association: Canberra:) 163-8
- Fraser, H. (1997a) "Dictionary pronunciation guides for English", *International Journal of Lexicography*. 10(3), 181-208
- Fraser, H. (1997b) "Phonology without tiers: Why the phonetic representation is not derived from the phonological representation", *Language Sciences*. 19(2), 101-137
- Fraser, H. (2000a) *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*, (Australian National Training Authority (ANTA Innovative Project): Canberra)
- Fraser, H.. (2000b) *Learn to Speak Clearly in English (CD-ROM for learners of English as a second language)*, (ANTA Innovative Project: Canberra) Available through Language Australia, Melbourne.
- Fraser, H. (2000c) "Literacy vs oral communication skills for ESL learners", *Literacy Link: Newsletter of the Australian Council for Adult Literacy*. 19(3), 4-6
- Fraser, H.. (2000d) "Pronunciation as Communication: Introducing a new CD-ROM for teaching ESL pronunciation", *Australian Language Matters*. 8(3)
- Fraser, H.. (in preparation) "A Teacher's Guide to Teaching Pronunciation to Learners of English as a Second Language" Available through Language Australia, Melbourne
- Löfqvist, A. (1997) "Theories and models of speech production", W. Hardcastle and J. Laver. *The Handbook of Phonetic Sciences*. (Blackwell: Oxford) 405-426
- Wierzbicka, A. (1990) "The meaning of color terms: semantics, culture and cognition", *Cognitive Linguistics*. 1(1), 99-150
- Winograd, T. and Flores, F. (1986) *Understanding Computers and Cognition: A New Foundation for Design* (Ablex, New Jersey)