Cauldwell, R. (2014). Listening and pronunciation need separate models of speech. In J. Levis & S. McCrocklin (Eds). *Proceedings of the 5th Pronunciation in Second Language Learning Teaching Conference* (pp.). Ames, IA: Iowa State University.

LISTENING AND PRONUNCIATION
NEED SEPARATE MODELS OF SPEECH

Richard Cauldwell, Speech in Action, Birmingham UK

Language learning pedagogy is currently dominated by the Careful Speech Model (CSM), key components of which are the citation forms of words, the relationship to grammatical categories (tone groups to clauses, intonation to questions, etc) and the rules of connected speech. However these aspects of the CSM – and particularly the rules of connected speech – are inadequate to characterize the unruliness, wildness, and massive reductions that occur in spontaneous speech. To teach listening more effectively we need a new model of speech – the Spontaneous Speech Model (SSM) – some of the components of which are outlined in this paper. I also argue: that we language teachers need a new mind-set which permits and accepts conflicts between what we teach in the CSM and what we teach in the SSM; that we need to take active steps to guard against our expert-listener tendency to hear full words where only massively reduced traces occur; that we need to let go of some of our favorite rules of connected speech.

The learning ghetto

A friend of mine – Tim – learnt Welsh as an adult. He told me of his pride in being able to have a conversation with his teacher in a non-classroom setting. But this pride turned to dismay when they were joined by a native speaker who then conversed with his teacher in spontaneous speech that Tim could not understand. From being confident and comfortable in Welsh, he suddenly felt excluded. He described this experience as feeling he was trapped in a 'language-learning ghetto'.

What Tim has identified is the fact that we ill-prepare language learners to handle the stream of spontaneous speech in real time. I believe that there is increasing realisation in the language teaching profession of the fact that we are not quite doing the right thing with the teaching of listening. Rost (2001: 13) writes that listening '…is still often considered a mysterious *black box* for which the best approach seems to be *more practice…*' We have yet to work systematically to open up the black box, and to go beyond the more-practice paradigm.

Separate goals

The way forward is pointed by Celce-Mucia et al (2010, p. 370) who state that the speech learners need to understand is 'much more varied and unpredictable than what they need to produce in order to be intelligible'. They conclude that for pronunciation and listening, 'the goals for mastery are different'.

My contention is that because the goals for mastery of pronunciation and listening are different, **we need two different models of speech** – one for pronunciation, which we already have, and an additional one (which we need to drag into existence) for listening. This model needs to be one that is an accurate representation of the realities of spontaneous speech. The trouble is that ELT is dominated by the model we already have, the Careful Speech Model in its British and North American variants (Cauldwell, 2013). The Careful Speech Model (henceforth CSM) is based on the citation forms of words, and consideration of how these citation forms would sound when joined

together and placed in clauses and sentences. Although it is useful for teaching pronunciation, this model of speech is not appropriate for teaching listening to everyday speech.

Need for a new model

In order to teach effectively towards a particular goal, we need to have a descriptive model of what we want learners to master, from which we can derive a syllabus, a metalanguage and teaching/learning activities. The CSM does not give us a model of spontaneous speech, it is a prescriptive model of orderly, rule-governed language. Its components can be compared to plants separately potted in a greenhouse (isolated citation forms), and of flower arrangements in a garden (rules of connected speech). However, for the purposes of teaching listening we need to prepare learners for the jungle of spontaneous speech, where the vegetation is crushed together in a messy and unruly manner which is quite unlike any orderly garden arrangement. Crucially, the mess and unruliness of spontaneous speech go way beyond what is generally allowed for in presentations of the rules of connected speech (such as those described in Brown, 2010; Celce Murcia et al, 2010; Roach, 2009).

So we need to embrace the mess and unruliness of everyday speech. And we will (for the purposes of listening) have to be prepared to let go of some of our favorite things to teach, to devise another model of speech, the Spontaneous Speech Model (henceforth SSM) which is a better description of the Jungle.

Many of the components of the CSM are expressed in terms of rules, and are exemplified with scripted acted speech. But acted speech is a completely different animal from spontaneous speech. The rules of the CSM can be viewed as hypotheses about what happens in spontaneous speech. But I suspect that most of them either have been refuted (as has happened for stress timing cf. Dauer, 1983, Roach, 1982; and for emotion in speech cf. Stibbard, 2001) or will be refuted. It is my belief that in spontaneous speech, the null hypothesis will be found to reign supreme. We should not expect that any of the CSM rules are true of spontaneous speech.

So one of the features of the spontaneous speech model (SSM) should be that it is unhooked from the expectations and rules of the CSM. I am not suggesting that we do away with the CSM – we should continue to allow it to perform the role that it has done for a long time, which is to act as a syllabus and a provider of metalanguage for the teaching of pronunciation.

Suggestions for components of the new model

There have been a number of scholars who have over the years pointed us in the direction of an SSM. The two editions of Brown's work (1977; 1990) and Shockey (2003) on the processes that operate on natural speech are landmarks, and in the mid nineteen-eighties Ur (1984, p. 17) suggested that we need to focus on words which are spoken 'quickly, in an unemphasized position in a sentence and juxtaposed with other words ...'. More recently, Field (2008, p. 196) suggests that the unit of decoding and understanding is neither the word nor the sentence, it is the intonation group: 'Words take their shapes from the intonation group as a whole and may not be identifiable until the whole group has been heard'.

A number of authors have favorite examples which they quote to show the extreme reductions that can happen in spontaneous speech. For example, Field (2003, p. 331) gives us 'narp meme' which is a reduced form of 'Know what I mean?'. And a number of the contributors to Brown (2010) present examples of extreme reductions embedded in their classroom activities. Their examples

point to a jungle-like speech world which lies beyond the rules of connected speech. These contributors include Jeff Stewart (ibid. p. 9), Michael Wilkins (p. 25), James Cassidy (p. 94), Mitsuyo Nakano (p. 224) and Arthur Nakano (p. 248). It is important that we do not regard such examples as interesting oddities, as exceptions to the rules which are familiar to us. They need to be put center stage, and treated as typical examples of what can and does happen in spontaneous speech: they need to be regarded as central to the model, not as outlying oddities. In phonetics and psycholinguistics there is an emerging literature on extreme reductions in which Johnson (2004) and Ernestus & Warner (2011) are key papers. This literature can provide important input to this model.

Escape from the graphic

We need also to escape the temptations and expectations of the forms of the written language, the graphic substance, which is the means by which most language teaching and learning takes place. We have to come to terms with the idea that students need to master the transient sound substance that is invisible and shaped by the speaker in real time. It is not a permanent graphic substance that is visible, it is not shaped by rule from the citation forms and the grammar.

The graphic substance can be seen, and it remains in view so that we can inspect it and edit it. And although there is some variation in the shape of words (different fonts, upper and lower case letters, some contractions) the words retain a close resemblance to their canonical shape.

Sound substance, on the other hand, is invisible. When it has happened it is gone – leaving traces in short-term memory which are then replaced by the traces for the new speech which is happening. And words (most often) do not occur on their own: they occur in rhythmic chunks – speech units – with some words emphasised and others not. And words are run together in a continuous acoustic blur where it is difficult to determine where one word ends and another begins. For learners the stream of speech is a continuously changing uncapturable fast-moving mush of sound substance.

A new mindset

In order to help learners master this sound substance we need to change some of our ideas and practices. So when we are working on the goal of listening, we need to accept that some of the things that we need to do may well be in conflict with our practices in the teaching of pronunciation, and in conflict with the CSM. This is because many of the rules and guidelines of the CSM are violated in spontaneous speech.

Drafting phenomena

Perhaps the most noticeable violation is the presence of drafting phenomena. In the CSM we advocate a tidy version of the language, akin to carefully edited written language. But spontaneous speech contains – as part of the sound substance – phenomena which would be edited out and would therefore be invisible in the written language. Because it is created and heard in real time, we hear the drafting phenomena: pauses for thought, filled pauses, references to speaker and hearer roles ('I mean', 'you know'), references to encoding difficulties ('How shall I put it'), appeals to shared understanding ('you know what I mean') and the changes of mind. These phenomena are crossed out in the written draft, and do not appear in the final version of a written text. But they are heard in the sound substance. In spontaneous speech, unlike in writing, you cannot stop for thought, erase or edit what you are saying without contributing to the sound substance – even if that

contribution is silence. The CSM deprecates such phenomena (and things such as 'like', 'and that kind of thing' 'and stuff') but this stuff occurs in spontaneous speech, and if we are to teach the sound substance of everyday speech, we should not pretend such phenomena don't exist. They can be kept out of the CSM, but they definitely need to be in the SSM.

Sound shapes and the blur gap

We need to make a deliberate effort to put the CSM out of our minds when we are teaching listening. This is because the CSM primes us to attend to spontaneous speech in a way which conforms to our expectations: its rules and guidelines become expectations, and that's what we hear. This leads to something I refer to as the blur gap (Cauldwell 2013, p. 17) which is the gap between what expert listeners believe they hear, and what is actually present in the sound substance. Countering the effects of the blur gap is not an easy thing to do. This is because, as an expert native speaker listener, I construct much of what I believe I hear. The sound substance which arrives at my ears consists of traces of words streamed in the mush of the acoustic blur of spontaneous speech. But I am largely unaware that these 'words' are made up of mushed blurred traces which can be very distant from the citation form. I will generally perceive a word-trace as something more substantial, as having a fuller soundshape closer to the citation form. Our ears play tricks on us. Our expertise with the CSM plays tricks on us. We do not have a good sense of what is happening in the stream of speech, so we need the help of technology to train our ears. One way of doing this is to use Audacity to cut up the stream of speech (Cauldwell 2013, p. 296) and to create audio concordances of words in speech units.

Letting go

As well as guarding against the blur gap, we need to let go of our attachment to the rules underlying (famous) examples of how speech works. For example word pairs such as 'might rain' and 'my train' which the CSM holds to sound different, may actually sound identical. Equally, different tenses may sound identical: the present tense 'they bought' might well sound identical to the past perfect 'they'd bought'. The underlying articulations may reflect the differences in grammar, but what is present in the sound substance may be identical. Thus distinctions which are very important in in the CSM may well be indistinguishable, or indeterminate, in spontaneous speech. This indeterminacy needs to be built in to the SSM. This particularly applies to negation – negative morphemes may be inaudible, so that there is no difference, in the sound substance between 'very legal' and 'very illegal'.

Conclusion

The late Professor John Sinclair of the University of Birmingham (UK) used to tell an anecdote about a German professor of English who complained to him that 'You teach us one English, and then mumble to each other in another English that we don't understand'. Like Tim, with Welsh, the German professor feels excluded from interactions s/he wants to be part of. We need to improve the teaching of listening by adopting a new goal, a new model (the SSM), and changing our mindset so that we don't exclude our learners from interactions that they really want to be part of.

ABOUT THE AUTHOR

Richard Cauldwell has taught English in France, Hong Kong, Japan, and the UK, where he worked at the University of Birmingham. Since 2001 he has published electronic and print materials for listening and pronunciation. Two of his publications have won British Council Innovations in ELT awards (ELTons): *Streaming Speech* (online course) in 2004 and *Cool Speech* (iPad app) in 2013. His *Phonology for Listening: Teaching the Stream of Speech* published in print in 2013 and as an eBook in 2014 has been shortlisted for three awards, and has received high praise in reviews from many sources, including *TESOL Quarterly*, and the *Journal of the International Phonetic Association*.

REFERENCES

Brown, G. (1990). *Listening to spoken English,* 2nd ed. Cambridge: Cambridge University Press.

Brown, J. D. (2012). *New ways in teaching connect speech.* [Ed.]. Alexandria, VA: TESOL International Association.

Cauldwell, R.T. (2013). *Phonology for listening: Teaching the stream of speech.* Birmingham: Speech in Action.

Celce-Murcia, M., Brinton, D. & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide.* New York: Cambridge University Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics, 39.* 253-260.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*(4), 325-334.

Field, J. (2008). *Listening in the language classroom.* Cambridge: Cambridge University Press.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama, & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis* (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (ed.) *Linguistic Controversies, Essays in Linguistic Theory and Practice,* 73-79. London: Edward Arnold.

Roach, P. (2009). *English phonetics and phonology: A practical course,* 4th ed. Cambridge: Cambridge University Press.

Rost, M. (2001). *Listening.* In R. Carter, and D. Nunan (eds.) *The Cambridge guide to teaching English to speakers of other languages* (pp. 7-13). Cambridge: Cambridge University Press.

Shockey, L. (2003). *Sound patterns of spoken English.* Oxford: Blackwell.

Stibbard, R. (2001). *Vocal expression of emotions in non-laboratory speech: An investigation of the Reading/Leeds emotion in speech project annotation data.* Doctoral dissertation: University of Reading.

Ur, P. (1984). *Teaching listening comprehension.* Cambridge: Cambridge University Press.