

Our Mumbling is Rumbled: *Streaming Speech: Listening and Pronunciation for Advanced Learners of English.*

Richard Cauldwell

Abstract.

Learners of English have long complained that we teach them one form of English pronunciation, but communicate with each other in another (fast mumbled) pronunciation which they can't understand. Learners are thus excluded from understanding real English. I argue that multimedia technology makes it possible to educate academics, teacher-trainers, teachers and learners about the true nature of our mumbled speech.

1 Introduction

Non-native speaker teachers of English are under pressure from their governments to improve their English language skills. Take, for example, Japan and Hong Kong: Japan is planning a massive re-training of English teachers to start in 2003 aiming to take the average English level of the 60,000 teachers of English in Junior and Senior High Schools up to TOEFL 550; in Hong Kong, teachers of English are under pressure to take language competence tests (Falvey & Coniam, 1999) as a condition for continuing in their roles as English teachers.

These are professional pressures of a severity that native speakers can only guess at. But some non-native speaker teachers feel an even greater pressure, the source of which lies in their own aspirations. In particular the aspiration to match native speaker abilities in speaking and listening. Many want to be better models of English for their students, both to improve their students' English, and to retain their students' respect, as these students encounter more 'real speech' from imported native-speakers-teachers, or from satellite broadcasting. These teachers are somewhat mystified by apparent attempts to block this 'match-the-native-speaker' aspiration by authors such as Keys (2000) and Jenkins (2001) whose recommendations for intelligible International English are often interpreted as deprecating the native speaker model.

In my experience, non-native-speaker teachers are very keen to have access to, and to learn more about, and match 'real speech' (however defined) so that they can answer their students' questions about the English of the satellite TV stations, and the internet – which textbooks do

not prepare them for. For decades, groups of non-native-speaker teachers have complained that the English we promote and teach is not the one we speak. They accuse us native-speakers of teaching one type of English, and of mumbling to each other in another type of English which outsiders cannot understand.

Over the past few years I have been developing, with groups of teachers of English from Japan, materials which help them with these pressures, and which share the secrets of native-speaker mumbling with them. The materials are designed to help non-native speaker teachers on three levels:

1. as learners of English who want to improve their pronunciation and listening;
2. as teachers of English who want to extend the range of methodological choices in the teaching of listening and pronunciation;
3. as applied linguists, observing and identifying the patterns and structures of spontaneous speech.

The materials comprise an electronic publication (a Windows CD-ROM) – *Streaming Speech: Listening and Pronunciation for Advanced Learners of English* (Cauldwell, 2002b). *Streaming Speech* is a publication in the tradition of Discourse Intonation which has seen David Brazil's *Pronunciation for Advanced Learners of English* (Brazil 1994), Barbara Bradford's *Intonation in Context* (Bradford 1988), and Martin Hewings's *Pronunciation Tasks* (Hewings 1993). But *Streaming Speech* differs from these publications in several important ways:

- it uses recordings of spontaneous (unscripted) speech as the source for listening exercises and as the model for pronunciation activities
- it focuses on fast speech, and shows the relationship between the fast original words, and the same words spoken slowly and clearly
- it makes extensive use of 'click-and-hear' technology – so that recorded extracts (whether they be one minute or one second in length) are available at the click of a mouse
- vowels, consonants, and clusters are practised in short stretches (normally less than two seconds) of speech – speech units –taken from the original recordings
- it provides a choice of voices (three male, three female) for pronunciation models
- wherever work is done on the pronunciation of segments, it is done in the context of a stretch of the stream of speech.

The first three items on this list make it impossible for the secrets of native speaker mumbling to stay hidden: banished is the attention to sanitized studio speech; gone is the adoration of the citation form; revealed in all its glory is the mess of spontaneous speech.

In Section 2 there follows an overview of *Streaming Speech*; and in Sections 3-5 there is a description of the three levels of help - language improvement, methodological innovation, applied linguistics.

2 Overview

Streaming Speech: Listening and Pronunciation for Advanced Learners of English is a ten-chapter electronic publication. It is aimed at people who aspire to be expert speakers and listeners to natural English speech. It aims to help users handle the fastest stretches of speech in their own listening, and to produce fluent stream-like flows of speech as necessary in communication. It is most appropriate for non-native speaker teachers of English, either in service, or in training.

The first eight chapters use recordings of four women and four men who work at the University of Birmingham in England, but who come from different parts of the UK (and in one case from the Republic of Ireland). Each of these chapters focuses on a set of speech sounds (Chapter 1, short vowels; Chapter 2, long vowels; etc), and on a feature of spontaneous speech (Chapter 1 elision, Chapter 2 rhythm, etc). Throughout, there is constant attention to the relationship between the fast norms of spontaneous speech and the slow speech of citation forms.

After a 'Welcome' screen, and a 'Goals' screen, each of the first eight chapters has six sections:

1. Listening: comprehension
2. Listening: focus
3. Discourse Features
4. Pronunciation: segments
5. Pronunciation: extended stretch
6. Review

The general pattern is to begin with listening, and to proceed to pronunciation via a 'Discourse Features' section. The Discourse Features section is pivotal to the structure: it is led into by the listening sections, it contains teacher education (applied linguistics) material, and it prepares for pronunciation work.

Chapters 9 & 10 do not follow this pattern. Chapter 9 gives the user the opportunity of selecting from six (three female, three male) voices – the speakers whose recordings are used in Chapters 1-6 – to work on the speech sounds of English. It offers the full range of speech sounds in the voices of six different speakers: it is a workshop in which users can choose a voice, choose a set of speech sounds to work with, and then listen, imitate, record and compare their own performances to those of the original recording.

Chapter 10 is directly aimed at teachers and teacher trainers. Entitled ‘A Window on Speech’ it provides training in observing identifying and explaining the structures and patterns of spontaneous speech (cf. Section 5 below).

3 The learning level

At this level, the materials aim to improve the listening and pronunciation skills of the user.

3.1 Listening

Section 1 of each chapter begins with a listening comprehension, where the user is first given a brief introduction to the speaker and the topic, and is then set a two or three questions to answer while listening. The recording is normally between one and two minutes long. The questions are answered by the user selecting the answers with clicks of the mouse, and these answers are recorded by the software. The exercise has an familiar design (introduce context, set questions, listen and answer questions) but the concepts behind the implementation of the design are new. The questions deliberately target ‘rich’ areas of the recording: rich in communicative content – they contain chunks of useful information – and rich in the features, patterns, and structures of spontaneous speech. Targetting the feature-rich, as well as the content-rich, moments of the recording is important for the design of the whole chapter: the feature-rich moments of the recording are used subsequently in Sections 3 (Discourse Features) and Section 4 (Pronunciation).

Section 2 ‘Focus’ of each chapter follows up on the questions asked in Section 1 (for the rationale cf Cauldwell 2002b). The focus section begins by letting the users know whether or not they have the right answer, and then – crucially – presents the extract from the recording which provides the evidence for the answer.

For example, in Chapter 1 one of the listening comprehension questions is ‘What type of evening classes did Corony set up?’ – the user is asked to answer in a multiple choice format. Having chosen an answer, the user then moves to a screen which contains the extract shown in Figure 1.

Figure 1.

```
014 // and we set up // 233
015 // a whole lot of evening classes // 258
016 // in // 153
017 // pottery and // 181
018 // woodwork and // 127
019 // drama and that kind of thing // 282
020 // erm // 047
021 // and i got very involved in those // 194
```

The notation is as follows: each line represents a speech-unit (in Brazil’s 1997 terminology, a tone-unit); the line begins with a number for the speech unit, which refers to its position in the transcript of the whole recording; the double slash (//) symbol signifies a speech unit boundary; between the boundary symbols are the words of the speech units and the line ends with the speed of the speech unit in words per minute.

Users click on a speaker ikon, and hear the extract: as each line plays, the relevant line of the transcript blushes red. Users can also click on any line, and hear that line in isolation.

The recorded evidence for answers is available in an accessible form; users are presented – via the click-and-hear technology – with real feature-rich (mumble-rich) speech in digestible chunks for inspection, both for listening improvement and as preparation for the work to come in the later sections.

3.2 Pronunciation

Sections 4 & 5 of each chapter focus on pronunciation. In each of the first eight chapters, Section 4 begins with a reminder, using words from the original recording, of the relationship between symbols and sounds (cf. Extract 2).

Figure 2 Reminder of the relationship between symbol and sound

	I	e	æ	ʌ	ɒ	ʊ
thing	✓					
very		✓				
that			✓			
up				✓		
obviously					✓	
putting						✓

Note: the symbols in the top row are those for the short vowels of English; the words in the first column contain the vowels given in the top row.

As everywhere in *Streaming Speech* the click-and-hear technology allows users to click on a word, and hear it spoken. Subsequent to the reminder, there is then a self-test, where the users test their knowledge of the the symbol/sound relationship with words extracted from the original recording.

Having done that test, and after having feedback, the user is presented with the table shown in Figure 3.

Figure 3 Table of short vowels and speech unit models

I	019. DRAMa and that kind of THING	282
	023. because i was still LI ving with	313
e	002. i was VE ry in VOLVED	164
	040. my OWN business as a TEX tile artist	227
æ	008. which i RAN	180
	032. i was a LE ady by THAT stage	197
ʌ	033. RUN ning	60
	037. MADE quite a bit of MO ney	330
ɒ	021. and I got very in VOLVED in those	194
	055. it was OB viously very PO pular	266
ʊ	063. CU shions and TA blecloths	143
	094. and AC tually PU ttling the things	233

Note: the symbols in the left hand column are those for the short vowels of English; the central column contains the sample speech units from the original recording, with the target sound shown in the syllable in bold upper-case letters; the right hand column shows the speed of the speech unit in words per minute. Syllables in upper-case are prominent syllables. The three digit number preceding each speech unit is the reference number of the speech unit in the original recording.

The user's task is to: first click on each line and listen to it a number of times; second to imitate the speech unit so that the target sound is

pronounced accurately in its position in the stream of the speech unit; third, when ready, to record their own version of each speech unit; fourth, compare their own version with the original, and then make an assessment of how well they have done.

In piloting, I discovered that students found matching the speed of the original more difficult than I had anticipated. So I have added advice, and added recorded examples (using my own voice), to demonstrate how to accelerate from slow pronunciation up to the speed of the original.

After to this work on specific sounds in speech units, users are presented with a related, but different task.

Figure 4 Streaming speech – practising with a sequence of speech units

032 // i was aLREAdy by THAT stage //	197	
033 // RUnning //	60	
034 // a SNACKbar //	139	
035 // THREE nights a WEEK //	230	
036 // WHICH //	69	
037 // MADE quite a bit of MOney //	330	

The pronunciation work ends with users listening, imitating, recording, comparing and assessing their pronunciation of sequences of speech units of the type shown in Extract 4. You will notice that these speech units are reasonably rich in short vowels, and that speech units 032, 033, & 037 appeared in the earlier pronunciation exercises (see Extract 3). This exercise is important for two main reasons: first, it brings meaning back into play – having worked earlier with isolated speech units, the user now works with a sequence of speech units which together make up a meaningful chunk of discourse; secondly, by practising a sequence of speech units, users get a feel of the variability of speech in extended stretches. This variability is demonstrated in Extract 4 by the difference in speeds of the speech units – varying speed like this is one of the ways in which speakers keep their listeners interested – and it is only by practising multi-speech unit stretches that this variability comes into play.

4 The methodological level

The methodological underpinnings of *Streaming Speech* are set out fairly extensively elsewhere (Cauldwell, 1996, 1998, 2002b), but one or two points are worth mentioning here.

4.1 Listening:

Most listening materials are concerned with *everything but* the sound substance, the mumbling, the acoustic blur of speech, which is central to the experience of listening. Most materials have the vices of being too attractive, too teacher and learner friendly, and too dependent on first language research (Cauldwell, 1998). They are over-concerned with topic, presentation, cognitive and metacognitive strategies, and the testing of extensive listening. They have very little post-test feedback: they are dominated by the misguided application of skills more appropriate for reading (Cauldwell 2002b).

Streaming Speech presents a syllabus for listening, an analysis of the ‘what’ of listening. It provides a chapter by chapter introduction to the patterns and structures of the stream of speech – it ‘grasps the nettle’ of fast speech (Cauldwell, 2002b). Central to the *Streaming Speech* approach is having transcript and original recording closely linked, so that after having done a listening task, there is feedback on the recorded evidence for the answer, and then a step by step introduction to the structures and patterns of the stream of speech.

Three considerations are crucial to the value of any listening lesson. First, students should be given activities which challenge and extend their abilities to hold in their minds traces (perceptual and meaningful) of the speech they have heard – without the support of the written language. This I refer to as ‘handling’ speech – and I cannot emphasise too much, this must be done with the sound substance without written support. So the second consideration is refraining from showing the students the transcript. The third consideration is showing students the transcript. Students should be given time to process, in their auditory memory, their perception and understanding of the recording. Students may feel lost or frustrated at this stage (cf. Cauldwell 2002b) – but this is fine, as long as they are not left in this state. After a certain amount of time handling speech (and only the teacher can decide the precise amount of frustration that the students can tolerate) then students should inspect the written transcript, and study the relationship between the sounds they heard, and the words and meaning they represent.

The crucial methodological stage is the attention to the sound substance of the recording: handling perceptual and meaningful traces without the transcript.

4.2 Pronunciation

Most pronunciation materials focus on the accurate articulation of individual sounds, and give separate treatment to intonation. In *Streaming Speech* all work on pronunciation is based on the original units of intonation (speech units) in which the individual sounds occur.

Because *Streaming Speech* is for advanced learners, there is no instructions on articulation of sounds – there is no ‘put your tongue here’ instruction. *Streaming Speech* works on the assumption that users have already had as much as they can manage of this type of instruction, and that they need more work on producing a natural stream-like flow. All the exercises are designed to help users produce speech as a fluent flow, with accurate differentiation of individual segments within the flow. They do so by imitating the meaningful utterances of native speakers.

5 *The applied linguist level*

Almost all phonology that is taught on teacher training courses is focused on pronunciation. It targets the inventory of vowel and consonant symbols that appear in dictionaries and vocabulary lists: and the major focus is on positions of the tongue, lips and teeth to produce speech sounds. The intonational apparatus is usually presented in later chapters, and usually consists of rules about the imagined reading aloud of isolated sentences: e.g. ‘a statement will end with a falling tone on the last lexical item’. As a result, the intonational apparatus is regarded as being something that appears at, or beyond, the far edge of the pronunciation syllabus. It is ‘the difficult bit’ in a late chapter that – if we teach slowly enough – will fall (thankfully) outside what can be taught in a term.

In *Streaming Speech* the intonational apparatus (speech units) is used from the very beginning as a presentational device. The components of intonation (speech units, tones, prominences, key) are the very stuff of speech – it is the medium in which vowels and consonants swim – and a major part of the philosophy of *Streaming Speech* that vowels and consonants should be presented in their natural environment. So, unlike conventional materials, the intonational apparatus is no longer at the far distant end of pronunciation, it is the vehicle by which all the recorded material is presented.

There is both a problem and an advantage to this approach. The problem is that very early on in *Streaming Speech*, it is necessary for users to start learning about speech units: what they are, what features they contain,

what features their boundaries have, etc. There is a learning load, to which users are likely to be unaccustomed: this is because listening materials are not normally burdened with applied linguist style explanations of the structure of speech. Thankfully, this problem is largely mitigated by the click-and-hear technology: users do not have to know what the intonational notation represents (though it helps, and explanations are available at the click of a mouse) because the original sound is just a click away – all users have to do is to click on the line of notation, and they hear the original.

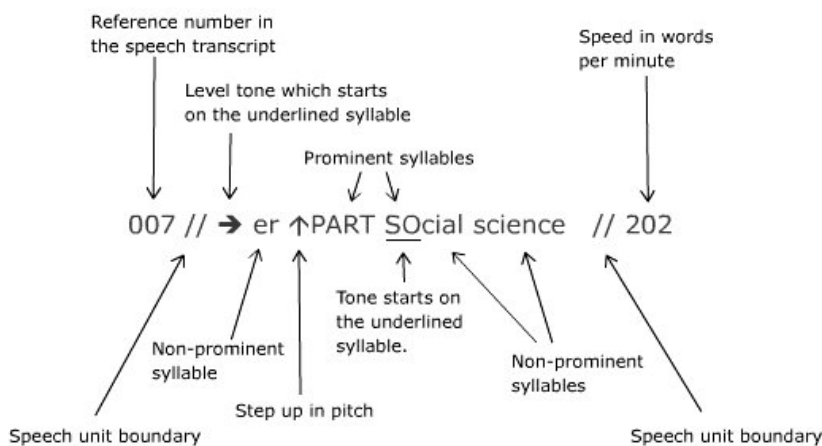
The advantage is that in introducing, step-by-step, the notion of the speech unit there actually emerges a syllabus for listening.

5.1 Discourse Features

The third section of Chapters 1-8 is entitled ‘Discourse Features’ and it is this section that provides an explanation of the features of speech units.

There is a step-by-step introduction to the concepts, notation, and auditory properties of the speech unit. Crucially, the click-and-hear technology allows the explanation of notation (shown in Figure 5) to go hand-in-hand with examples from the original recordings. There is constant, consistent, association between sound and transcription: having this ever-present association is vital to the whole publication.

Figure 5 The Notation of the speech units



The syllabus covered in each of the Discourse Features sections is listed in Table 1.

Table 1 Contents of the Discourse Features sections

1	Elisions	Varying speed to keep listeners' attention
2	Prominences	Varying rhythms to keep listeners' attention
3	Tones Falling & Level	Pausing on words to give planning time
4	Tones: Rising	Tones on questions
5	Speech units	Varying speech unit size; words prominent & non-prominent
6	Stress shift	Stress patterns of words in speech units
7	Key: high & low	Ending questions, beginning answers
8	[All the above]	Using all the strategies to tell an anecdote

Table 1 shows topic in column 2 and the speech strategy to which it is related in column 3. Notice that the word 'varying' appears several times in this third column. It is the variability of speech that is most striking when viewed via the click-and-hear technology of *Streaming Speech*. Variability is the key to listenability (maintaining the listener's interest) which you do by continually varying, in non-predictable ways, the shape of the stream of speech. If, perversely, you construct your speech according to textbook rules, you will very soon produce predictably patterned, distractingly repetitious speech. (Fortunately this is quite difficult to do without rehearsal). So many of the observations that emerge from the Discourse Features sections conflict with textbook rules: such as those relating to the relationship of speech units to clauses, to question intonation, to 'stress-timing' (cf also Cauldwell 2002a).

5.2 A window on speech

Chapter 10, the last chapter in *Streaming Speech* is aimed directly at teachers: it contains training in recognising speech units. It is in three parts: part 1 contains instruction in identifying the five different sizes of speech unit in any stretch of speech; part 2 deals with the five tones of English both on monosyllables, and polysyllabic stretches; part 3 contains four transcription exercises – with answers. The aim of this chapter is to help teachers identify speech units in other recorded materials, so they can understand for themselves, and can explain to their pupils, what is going on in the recording.

6 Conclusion – our mumbling is rumbled

Streaming Speech is the outcome of over ten years of working with non-native speaker teachers of English. The three-level approach – skills improvement, methodological suggestion, and applied linguistic study –, arose out of a set of needs from a particular group of trainees, but it is proving popular with groups of teachers other than those originally taught

(Japanese teachers of English). It seems to me that there is a world-wide demand for materials such as this which make natural spontaneous speech accessible, and understandable – both in the listening comprehension sense, and in the applied linguistic sense of understanding how it works. With this use of technology, our mumbling is rumbled – the secrets of what native speakers do at speed in spontaneous speech are laid open to the world.

Streaming Speech is designed to be used in self-study mode but it has also been used in whole class situations, in which the teacher controls the computer, and learners use hardcopy workbooks to annotate, and view the scripts. A workbook will be published later in 2003.

Sometimes people associate CD-ROMs with games, exercises and supplementary materials – *Streaming Speech* is different, it is a complete course of pronunciation, listening, and in training in observing and identifying the structures and patterns of natural speech. No mere written text can hope to show the benefits of the click-and-hear technology. Fortunately, it is possible to try out Streaming Speech on the web at <http://www.fab24.net/examples/streamingspeech.htm> (You will need Internet Explorer 5.5 or later, and Macromedia Flash Player. You will need to say ‘yes’ when invited to download a sound-recorder.).

- Bradford, B. (1988). *Intonation in Context*. Cambridge: Cambridge University Press.
- Brazil, D. (1994). *Pronunciation for Advanced Learners of English*. Cambridge: Cambridge University Press.
- Brazil, D. (1997). *The Communicative Value of Intonation in English*. [2nd Edition]. Cambridge: Cambridge University Press.
- Cauldwell, R.T. (1996). Direct Encounters with fast speech on CD Audio to teach listening. *System*, 24/4, 521-528. (ISSN 0346-251X).
- Cauldwell, R.T. (1998) Faith, hope, and charity: the vices of listening comprehension. *The Language Teacher* July 1998
- Cauldwell, R.T. (2002a) The Functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Applied Language Studies*, 2/1 [Paper published on the web <http://www.solki.jyu.fi/apples/>]
- Cauldwell, R.T. (2002b) Grasping the nettle: The importance of perception work in listening comprehension. [Paper published on the web http://www.developingteachers.com/articles_tchtraining/perception1_richard.htm]
- Cauldwell, R.T. (2002c). *Streaming Speech: Listening and Pronunciation for Advanced Learners of English*. Birmingham: speechinaction.
- Falvey, P. & Coniam, D. (1999) Language Benchmarks for Lower-Secondary Teachers of English: Findings and emerging issues. [Conference paper delivered at 'Issues In Language Proficiency Assessment for Teachers (LPAT) (Previously known as English Language Benchmarking) Proceedings of a Colloquium hosted by the Department of English at the Hong Kong Institute of Education, on Wednesday, 29 September 1999]
- Hewings, M. (1993). *Pronunciation Tasks*. Cambridge: Cambridge University Press.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Keys, K.J. (1999). Making the right noises: pronunciation and teacher training in Brazil. *Speak Out*, 24, 40-44.