

Bricking up and Streaming down: two approaches to naturalness in pronunciation materials

Richard Cauldwell, SPEECHINACTION

1 Introduction

All pronunciation materials aim to get learners to the point where they can speak with natural fluency. Most materials address this goal by building from the ground up. Starting with citation forms ('bricks'), which are built up into larger groups, then sentences. The rules and guidance in this approach come from an honoured tradition of consideration of how words and sentences should be uttered - in other words, naturalness is constructed.

Another approach which has for some time been technologically possible is to capture units of the rough and tumble of everyday speech (sections of the stream) and to break it down into learnable chunks. The advantage of the second (streaming down) approach is that naturalness does not have to be constructed - but it brings with it problems of idiosyncrasy, context-boundness, and the dangers of over-generalising from the single - possibly unique - instance. This paper will describe (as far as the printed word allows) how recently published materials attempt to get around these problems by using intonational categories (tone-units/speech-units) as a window on speech.

2 Bricking up

First of all, a confession. I'm not trained in the kind of analysis presented in this section. All I wish to do here is to remind you of the type of analysis that is typically done in the approach I refer to as 'bricking up'.

Consider the sentence in example 1, and how we might devise rules for its prosodic pronunciation:

ex. 1. This is one I'm going to be looking at in slightly more detail in fact.

A bricking-up approach might first consider which words are the weak forms, and which not, and would probably advise that some words were strong forms, and would receive stresses as shown in example 2 with a syllable in upper case:

ex. 2. this is ONE I'm GOing to be LOOKing at in SLIGHTly MORE DEtail in FACT

We might then advise that there is another level of stress above word stress - 'phrasal stress' - which would involve assigning stresses to 'going', 'looking', 'slightly' and 'detail'. In ex. 3 this level of stress is represented in bold:

ex. 3. . this is ONE I'm **GO**ing to be **LOOK**ing at in **SLIGHT**ly MORE **DE**tail in FACT

As the final stage, we might advise that the last content word would receive nuclear stress, in ex. 4 this is represented by the underlining on the first syllable of 'detail':

ex. 4. . this is ONE I'm **GO**ing to be **LOOK**ing at in **SLIGHT**ly MORE **DE**tail in FACT

Having done this, we would get the kind of stress assignment as shown in Table 1.

4							X	
3			X	X	X		X	
2		X	X	X	X	X	X	X
	this is	ONE	i'm GOing	to be LOOKing at	in SLIGHTly	MORE	DEtail	in FACT

Table 1. Different levels of stress assignment

Note that at the end of this procedure that the soundshapes of the 'non-weak' forms survive, even though they are not assigned stress at level 4. This kind of presentation is useful in that it preserves sound-shape information at each level, and the information builds up layer upon layer. Another usefulness of this approach is that it makes it possible to formulate pedagogic rules for language teaching.

3 Streaming down

In the streaming down approach, we work with recordings of used (not invented) language and use a simple set of conventions to represent what is heard, not what is expected or what ought to happen. The only level of stress that is represented is that of 'prominence': the speaker's choice of what to highlight for the purposes of communication at the moment of speaking (prominences are indicated by upper-case letters). It is not the practice in this approach to assign other levels of stress – word-stress is assumed to be part of the sound-shape of the word in the lexicon, a given, and not amenable to speaker choice, and this system of transcription records high level contextually related choices, not features which are linguistic 'givens'. The problem with this approach is that any analysis reflects the messiness of the evidence of the recording – and it therefore lends itself much less to the formulation of pedagogic rules and guidance. As it happens, though, our example sentence is unusually tidy in its coincidence of syntax and prosody: it is rare that such a long sentence co-occurs with a speech unit so neatly. But its central section has an instructive messiness which we will come to shortly.

The sentence was recorded at a university seminar in the mid 1990s (Cauldwell 2002 chapter 5). It is not a scripted sentence, the speaker used notes, not a script. In the 'streaming down' transcription conventions that I use, (derived from Brazil 1997) the sentence is represented as in ex. 5

ex. 5 // this is ONE i'm going to be looking at in slightly more DEtail in fact // 400

The location of the tone (in this case falling) is indicated by the underlining on 'detail'. It is important to note, for the purposes of the later discussion, that this speech unit was spoken at 400 words per minute.

This form of transcription tells us that at this moment in the lecture the speaker chose to highlight the words 'one' and 'detail' and did not highlight the other words. But it tells us nothing about the fate of the other words. However, those experienced in this mode of transcription will be surprised by the number of words in between the two prominences, and would immediately want to attend to what has happened to them.

As this is a printed publication, I cannot play you the recording – but I invite you to look

at Figure 1 to get some idea of the fate of the words. Again you will have to forgive me for the contents of Figure 1 – it is a screen-shot from Cauldwell (2002) in which I use a set of phonemic symbols for pedagogic purposes, and occasionally I try to represent - using just phonemic symbols – phenomena that should strictly require a much fuller use of phonetic symbols.

Figure 1 shows the words of the sentence in the left-hand column; the middle column shows how someone with Geoff’s accent (note the pronunciation of ‘one’) would say the words in slow careful speech; the right-hand column shows a phonemic representation of the fate of the words.

	slow careful	fast
this is	ðis ɪz	?ɪsɪz
one	wɒn	wɒn
i’m going	aɪm ɡeɪŋ	aɪŋgeʊ
to be looking at	tu: bi: lʊkɪŋ æt	lʊŋjæt
in slightly more	ɪn slɑɪtli mɔ:	ɪŋslɑɪ?lɪmu:
detail	dɪteɪl	dɪteɪl
in fact	ɪn fækt	ɪnfækt

Figure 1. Geoff’s sentence

At the fairly extreme speed at which Geoff was speaking – 400 words per minute – some extreme distortions happen to the words: note in particular that the words ‘to be’ have no existence that I can detect in the speech signal, and that the word ‘more’ is close to a very short sharp (un-cow-like) ‘moo’.

The transcription for ex 5. above can be presented as a five-part structure, as shown in Figure 2.

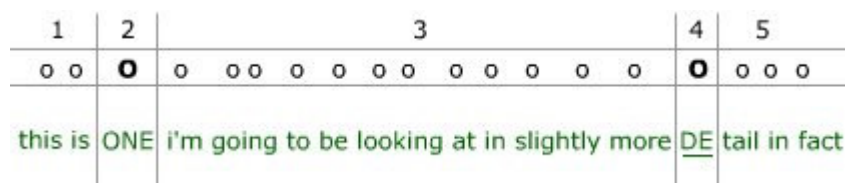


Figure 2. Five-part structure

You will notice, if you compare Figure 1 and Figure 2, that it is part 3 of the structure that is most destructive of the sound-shapes of words. Because of speaker highlighting choices, parts 2 and 4 are clear, and because this speech unit is pre-pausal, part 5 features a slowing down and final lengthening which results in the syllables ‘tail in fact’ surviving.

Immediately the question arises: do all parts 3 in similar five-part speech units feature such sound-shape destructions? And are there generalisable rules of fast speech that can be derived from inspection of this and other examples?

Strictly speaking the answer is ‘I don’t know’ I haven’t looked at sufficient number of

speech units to come to any conclusions. But Table 2 presents other examples worth discussing.

1	2	3	4	5
this is	ONE	produced by	PRIEST	ley
a	COUN	cil called the student	FASH	ion council

Table 2. Five-part structure

What is of interest is the fate of the words 'produced' and 'student' – both of them become virtually monosyllabic, with the /d/ suffering a destructive fate which it is beyond my competence to describe. What is also of interest is that the speed of the speech units is at around 200 words per minute – so even at relatively low speeds, the destructive effect of part 3 is observable. What I would be happy to assert that the notion of the five-part speech unit provides a window on speech which is proving to be extremely useful in making evidence-based observations of speech which are useful at research, teacher-training, and classroom levels, but cannot predict the exact nature of the sound shape destructions

4 A window on speech

In the previous section, I have focussed on five part speech units with two prominences: there are other sizes of speech unit, five in all, which comprise what is termed 'A window on speech' (see chapters 10 of both Cauldwell 2002 & 2005). They are patterns of the sound substance (the stream) of speech which record choices speakers make in highlighting what they say in context. Together with Macromedia Flash™ technology they comprise an observational framework that provides insights into the evidence of the fate of the sound-shapes of words. They help capture the naturalness of 'deaccenting' in speech which is useful both in listening pedagogy (for which it is necessary to acclimatise learners to what happens at 400 words per minute) and in providing samples of natural speech for pronunciation purposes – albeit at lower speeds, and with less extreme relationships between prominence and the soundshape of words.

I believe that the five shapes of speech unit comprise a useful set of significant sound-shapes of English for the capturing of prosodic phenomena.

5 References

Brazil, David (1997) *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

Cauldwell, Richard (2002) *Streaming Speech: Listening and Pronunciation for Advanced Learners of English – British/Irish version*. [Online publication at <http://www.speechinaction.com>] Birmingham: speechinaction.

Cauldwell, Richard (2005) *Streaming Speech: Listening and Pronunciation for Advanced Learners of English – American/Canadian version*. [Online publication at <http://www.speechinaction.com>] Birmingham: speechinaction.