

WHERE DID THE ANGER GO? THE ROLE OF CONTEXT IN INTERPRETING EMOTION IN SPEECH.

Richard T. Cauldwell

EISU, Department of English
The University of Birmingham

ABSTRACT

Context has greater explanatory force regarding the perception of emotions in speech than any causative association we may attempt to make between phonetic features and psychological labels. Informants, when asked to judge whether a speaker sounds angry or not, make different decisions according to whether they hear utterances in-isolation or in-context. There was a strong tendency for 'anger' to be heard in the isolated extracts, but not in the 'in-context' forms.

1. LEXIS IN CONTEXT

The way that a sequence of words is uttered can cue attitudinal meanings which lie beyond the explanatory power of lexicogrammatical, and discourse-structure tools. Thus a question such as 'What do you mean?' can be uttered in such a way as to cue the presence of attitudinal meanings as diverse as *displeasure* or *delight*. This paper is concerned with perceptions of such attitudinal meanings when sections of spontaneous speech, are played to a group of informants, in-isolation and in-context. I argue that traditional tables of associations between tones and attitudes (henceforth 'attitude tables') are misleading; and I want to demonstrate that the meaning potential attributed to different ways of saying isolated utterances can be neutralised by contextual factors.

2. PROSODY, ATTITUDES AND EMOTIONS

For decades, one of the goals of intonation studies has been to find systematic links between elements of prosody and emotional/attitudinal (henceforth, *attitudinal*) meanings. As time progressed, and studies have failed to identify reliable links, there has been an increasing recognition of the importance of contextual factors, which can be illustrated by tracking the work of Crystal (1969; 1975; 1995).

Crystal (1969) conducted an experiment to find the relationship between items of non-segmental phonology such as falling and rising tones, and attitudes such as 'excited', 'dismayed', 'haughty', 'puzzled', 'angry' (these are just five of twenty such

labels). Six informants were asked to read aloud three isolated sentences (e.g. 'There's that little black dog in the garden again') in the tone of voice they associated with each of the twenty labels 'after imagining themselves in an appropriate situation' (1969:297). A sample of the findings can be seen in Table 1.

Tones	Labels
fall	conspiratorial, angry, matter of fact, vexed, impatient, satisfied, <i>grim</i> , irritated
rise-fall	excited, <i>pleased</i> , <i>dismayed</i> , <i>haughty</i> , amused
fall-rise	Worried, <i>pleased</i> , <i>dismayed</i> , disapproving, apologetic, <i>grim</i>
rise	Puzzled, <i>bored</i> , questioning
level	<i>bored</i>

Table 1: Associations of tones with attitudinal labels (Crystal, 1969:305)

Column 1 shows tones which Crystal found to be 'important characterising features' of the attitudes listed in Column 2. Note that there are no exclusive relationships between tone and attitude: the tones (except the level tone) are associated with more than one attitude; and that a number of attitudes (italicised in the table) are associated with more than one tone. These cross-associations were characteristic of other parts of the experiment.

This is largely a negative result, because it does not allow one to say that a given attitude is caused by a particular tone. Crystal acknowledged that 'a given non-segmental feature rarely has a single semantic field' and went on to suggest that 'it may be more useful to talk of contours having 'primary' functions in specific situations...' (p. 306). It is difficult to know precisely what Crystal means by this, but he seems to be suggesting that combinations of non-segmental features ('contours') had a role to play in 'specific situations'.

Six years later, Crystal attempted to associate contours (tones, plus 'drops' or 'boosters') with contextual information: part of this attempt is shown in Table 2 (derived from Crystal, 1975), however no 'specific situation' is mentioned.

Contours	Syntactic Environment	Meanings
Falling tone + low drop	final tonic in S	<i>attitudinally neutral</i>
	non-final tonic in Sentence	<i>personal definitiveness</i> abrupt, insistent <i>unsociability</i> cool, irritated, rude
Rising tone + Low drop	non-final tonic	<i>attitudinally neutral</i>)
	final tonic of Sentence	<i>personal inconclusiveness</i> non-committal, unaggressive, polite, respectful <i>social openness</i> casual, friendly persuasive, warning, grim (<u>with appropriate kinesics</u>)
Falling tone + high booster	in any position	<i>definite emotional commitment</i> emphasis, surprise, warmth, <u>depending on kinesics</u>
Rising tone + high booster	in any position	<i>definite emotional enquiry</i> query, puzzlement, surprise
Rising-falling	in any position	<i>definite outcome</i> impressed, satisfied, smug, or the reverse, <u>depending on kinesic accompaniment</u>

Table 2: Contours and their attitudinal meanings (derived from Crystal 1975, p. 38)

The contours (Column 1) consist of combinations of different types of non-segmental features: tones, plus ‘low drops’ and ‘high boosters’. The meanings (Column 3) are of two types: general meanings – given in italics – and specific meanings which apply in particular contexts. Thus reading across the table, the falling tone plus low drop (i.e. a fall to the bottom of the speaker’s pitch range), on the final tonic syllable of a sentence is ‘attitudinally neutral’; whereas non-finally it can mean ‘personal definitiveness’ or ‘unsociability’, or a specific meaning such as ‘abrupt’ or ‘irritated’. Note that context appears in the table in a number of different guises: Column 2 specifies a grammatical context; Column 3 twice indicates that the choice of a specific meaning (e.g. ‘friendly’ or ‘grim’) is dependent on the kinesics that occur in a particular context; and the very fact that there are different (and conflicting) labels indicates that the meanings are to a large extent context-dependent.

It is clear that contextual factors are important, but it is not clear which of the two sets of factors (contextual and prosodic) is the more important or powerful in communicating attitudinal meanings. However, there is an indication, in the last row of the table, that contextual factors (‘kinesics’) might be all important: Crystal asserts that a rising-falling tone can mean ‘impressed’, ‘satisfied’ or ‘smug’ ‘or the reverse, depending on the kinesic accompaniment’: the non-prosodic situational factors are sufficiently powerful to reverse meanings.

Twenty years later, Crystal (1995) provides a ‘specific situation’, inviting readers to imagine the word ‘Yes’ being spoken in answer to the question ‘Will you marry me’. Table 3 gives an extract of the inventory of tones.

Tone	Meaning
full fall	emotionally involved: surprise, excitement, irritation
full rise	emotionally involved: disbelief, shock
fall-rise	<u>a positive face</u> : encouragement, urgency <u>a straight or negative face</u> : uncertainty, doubt, tentativeness

Table 3: Attitudinal meanings on ‘Yes’ (derived from Crystal 1995: 248)

Note: (a) the terms ‘positive face’ and ‘straight or negative face’ are meant to refer to encouraging and non-encouraging facial expressions; (b) by ‘full’ Crystal means that the fall and the rise extend the full range of the speaker’s voice.

Even in this very specific context, it proves difficult to provide exclusive associations: similar attitudes occur with different tones (‘emotional involvement’ and ‘surprise’/‘shock’ occur with both the fall and the rise); and different attitudes are associated with the same tone (‘urgency’ and ‘tentativeness’ with the fall-rise). Again, the question arises of whether the power in conveying attitudinal meaning lies in the non-prosodic situational factors (a ‘positive ... or negative face’), or in the prosodic (tone) choices.

3. SHORT ISOLATED STRETCHES

Tables that purport to show links between tones and attitudes have a feature that requires comment at this point. Everyday speech rarely involves short isolated utterances: we typically engage in interactions of many utterances containing many tone-units. However, attitude tables focus only on single tone-units. The issue of how such isolated utterances, with their attendant prosodic features and attitudinal meanings, would fare in larger stretches of discourse is rarely addressed. This issue is one which I address in the two studies which are reported below.

4. STUDY 1

Two wh-questions (Extracts 1 & 2 below) were digitally extracted from a recording of a dinner-table conversation. The notation follows the conventions of Brazil (1994): they have falling tones which start on the underlined syllables. According to Crystal and Halliday they should sound attitudinally neutral: Crystal (Table 2 above) asserts that falling tone plus a low drop on a sentence final tonic syllable is ‘attitudinally neutral’; Halliday (1994, p. 305) tells us that the falling tone is the ‘unmarked’ choice for a wh-question.

[Click on the extracts to hear them.]

Extract 1 WHAT do you MEAN

Extract 2 WHEN

These extracts were played to thirty informants (all postgraduate students of Applied Linguistics: 13 native, and 17 non-native speakers, 7 nationalities) who were asked to make judgments about these extracts. No information was given about the context from which these extracts came. The judgement requested was a binary one: is the speaker angry/irritated or not angry/irritated?. The extracts were played three times, and students recorded their judgements by ticking boxes in a table.

If you are reading this as a word document, double-click on the icons to hear the extracts. If you are reading this in Acrobat, you can click on each example, and hear it.

I then asked them to listen to the utterance with contextual information supplied: that this was a parent-child interaction during a dinner table conversation involving Matthew aged three and a half, Mum in her late thirties, brother Richard aged thirteen and Dad in his early forties. For this stage in the procedure, the informants had a script of the interaction in front of them (Extract 3, below) so they could more easily locate the utterances in the recording. They heard the utterances situated in a 45 second extract twice. They were then asked to make the same judgement (irritated/angry or not) as they had for the utterances in isolation. The original extracts 1 & 2 feature as tone-units 18 and 21 in Extract 3. [Click the extract to hear it.]

Extract 3

1	Matt	mi↑AO <u>BUM</u> ma
2	Matt	[chuckle]
3	Matt	...i mean...
4	Matt	<u>HEY</u>
5	Matt	...HEY...
6	Matt	<u>WHEN</u> it
7	Matt	...when...
8	Matt	when when <u>WHEN</u> it when it <u>SNOWS</u>
9	Matt	...we can...
10	Matt	<u>DRIVE</u>
11	Mum	AND/what were you going to <u>SAY</u>
12	Matt	<u>YEAH</u>
13	Matt/Rich	<u>YEAH</u> yeah/ <u>DAD</u>
14	Matt	and <u>OUR</u> car
15	Matt	got <u>STUCK</u>
16	Dad	...our c wha wha...
17	Dad	... <u>WHAT</u> ...
18	Dad	WHAT do you MEAN
19	Matt	<u>OUR</u> car got <u>STUCK</u>
20	Matt	in the <u>SNOW</u>
21	Dad	WHEN

Note: Extract 3 was in fact longer than this, and informants were asked to make more judgments than are reported in this paper. Extract 3 is shortened, and reports of the other judgments omitted, for reasons of space.

The results are shown in Table 4.

	Isolated <i>Irritated/Angry</i>			In context <i>Irritated/Angry</i>			Swing
	Yes	?	No	Yes	?	No	
Ex. 1	66	7	27	10	7	83	56
Ex. 2	57	10	33	10	0	90	47

Table 4: Percentage of informants judging the extracts to be *Irritated/Angry* or *Not Irritated/Angry*. The Swing is calculated by subtracting percentage of *in context* 'yes' from percentage of *isolated* 'yes'.

Table 4 shows a marked swing between the judgments made of the wh-questions in the in-isolation Extracts 1 & 2 and the judgments of the same wh-questions heard in-context in Extract 3. In the case of in-isolation Extract 1 – ‘What do you mean?’ – 66% of informants judged it to be indicating *anger* and 27% did not. On hearing it in-context, the percentage of informants making *anger* judgments dropped dramatically by 56%; and there was a corresponding rise in the percentage of informants making *neutral* judgments. Similar figures apply to Extract 2.

4.1 STUDY 1:DISCUSSION

This first study showed that there was large majority of initial judgements for *anger* when the extracts were heard in isolation, but that there was a large swing towards *neutral* when the extracts were heard in context.

Where did the anger go? The answer is that anger went nowhere, it was never ‘in’ the extracts – it was neither in the wording nor in the prosody. *Anger* judgments arose because the study required informants to make a forced choice, and their perceptions of the way the in-isolation Extracts 1 & 2 were uttered matched better their model for *anger*, than for *neutral*. Why did they do so?

They did so because they have characteristics of speed and tenseness which, heard in isolation, mark them as outside the ‘normal range’ for an average adult male voice: they are perceived as having a ‘rarity value’ (Cauldwell, 1997) appropriate to a cueing of *anger*.

However, the ‘normal range’ (which we can informally characterise as ‘not cueing attitudinal/emotional meaning’) itself varies with context. The normal range varies with the time of day, the nature of accompanying activities, and who the addressees are. Hearers in a particular context instantaneously compensate for these variations in the normal range, and will only attend to marked excursions from the current setting for normal to evaluate whether or not these excursions are cueing the presence of attitudinal/emotional meanings. When hearers have little or no contextual information, when they have no prior experience of the speaker’s voice, they do not have sufficient time to attune to the speaker’s current normal range, and are thus likely to make judgements based on their past experience of similar voices in familiar contexts.

Thus (I speculate) informants’ perceptions of in-isolation Extracts 1 & 2 did not match their experience of normal ranges for adult male speakers, so (having to make a forced choice) they judged the Extracts to be conveying *anger*. When they heard the contextualised versions, however, informants had sufficient information concerning the current normal setting for Dad’s voice for them to realise that the prosody of the wh-questions lay well within the current setting of ‘normal’ for this speaker at this moment in the dinner-table conversation.

An interesting point to note is that the swing away from *anger* was greater for native speakers than it was for non-native speakers. This is shown in Table 5.

Extract		Range	Swing
1	Native Speakers	77-0	77%
1	Non -Native Speakers	59-18	41%
2	Native Speakers	62-0	62%
2	Non -Native Speakers	53-18	35%

Table 5: Comparison of native and non-native swings away from *Anger/Irritation* judgments (having heard Extracts 1 & 2 first).

Native speaker judgments were more extreme than non-native speaker judgments: they went from 77% and 62% *anger* to zero. These swings away from *anger* were respectively 36 and 27 percentage points higher than those for non-native speakers.

It would seem entirely reasonable to suppose that native speakers are more sensitive than non-native speakers to the ‘normal ranges’ for an adult male speaking English, and that they would be more sensitive to variations from the normal which would have rarity value. This would explain why native speakers were more extreme in their judgments. However, see Section 5.1 below for a cautionary note on this issue.

This study shows the importance of factoring in contextual knowledge into our statements of associations between prosody and attitudes. And the contextual parameters are not simply situational, and kinesic, as Crystal suggests. They also need to be prosodic: we need to factor in information concerning the current settings for normal for a particular person’s voice at a specific moment in a specific situation.

Having completed this study I was convinced that contextual factors were all powerful. I was sure that there would be no such swing if I conducted another study using the same data, but changing the order of the extracts – so that Extract 3 (the contextualised version) was heard first. I felt that informants’ memories of the contextualised version – having given them the information of the current settings for normal for the speaker – would cause them to judge the isolated versions (Extracts 1 & 2) ‘neutral’. I therefore conducted a second study.

5 STUDY 2

There was a separate group of informants for the second study consisting of 47 lecturers/professors and postgraduate students of Applied Linguistics, 15 nationalities, 19 native, 28 non-native speakers at the Universities of Birmingham and Reading. I first played the in-context extract (Extract 3), and asked informants to make the same *Anger* judgments as in the previous study. I then played the isolated wh-questions (Extracts 1 & 2), and asked the informants to make the *Anger* judgments once again. The results are shown in Table 6.

	In context <i>Irritated/Angry</i>			Isolated <i>Irritated/Angry</i>			Swing
	Yes	?	No	Yes	?	No	
Ex. 1	13	0	87	72	0	28	59
Ex. 2	21	0	79	68	0	32	47

Table 6: Percentage of subjects (N= 47) judging the extracts to be *Irritated/Angry* or *Not Irritated/Angry*. Informants heard the in-context extract (Extract 3) first. The swing is calculated by subtracting percentage of in-context ‘yes’ from percentage of in-isolation ‘yes’.

Table 6 shows that, against expectations, there was a marked swing of almost exactly the same order, in percentage terms, as in the first study. As informants heard the in-context version first and the isolated forms second, the swing this time was towards *anger*. In other words, in this, as in the previous study there was a strong association of *anger* with the isolated extracts, and a strong association of *neutral* with the in-context versions.

5.1 STUDY 2:DISCUSSION

The results of the second study seem to refute the explanation given for the swings of the first study. Having heard the in-context version, the informants had the information they needed concerning the current settings for ‘normal’ for Dad’s voice: they were attuned to the normal parameters for this voice in this context, but they still swung, as had the first group of informants, towards *anger* on hearing the isolated extracts.

However I believe there is an explanation for this unexpected result which is consistent with the suggestions contained in the discussion of Study 1.

Hearers have to be continually adjusting to the current settings for normal and they do so at great speed, and with great facility. Their knowledge that they are hearing the same speaker in-isolation as they had heard in-context does not over-ride their capacity to adjust continually to the current settings for normal. On moving from judging the in-context versions to judging the isolated versions their capacity to accommodate to the current settings for normal remains operative. Their memory of in-context judgments does not dominate the subliminal capacity to accommodate to speech norms. It could be, therefore, that the ‘no swing’ expectations that I had before engaging on Study 2 were unknowingly relying on a fallacy: the expectation of the suspension/fixing of a process (accommodating to the current setting for normal) that in fact never stops.

Study 1 indicated that native speakers tended to make more extreme judgments: the swing from *anger* to *neutral* was much greater than it was for non-native speakers. As can be seen from Table 7 this tendency was not repeated in Study 2.

Extract		Range	Swing
1	Native Speakers	16-79	63%
1	Non -Native Speakers	11-68	57%
2	Native Speakers	26-53	27%
2	Non -Native Speakers	18-79	61%

Table 7: Comparison of native and non-native swings towards *Anger* judgments in Study 2.

Table 7 shows that for Extract 1 ('What do you mean?') the native speaker swing was only 6 percentage points greater than the non-native speaker swing (as against 36 percentage points greater in Study 1).

Even more interestingly, for Extract 2 ('When?'), non-native speakers had a swing towards *anger* which was much greater than for native speakers. Whereas in Study 1 the native speaker swing was 27 percentage points greater than the non-native speaker swing, in this study the non-native speaker swing is greater by 34 percentage points, making the swing more than twice as large as the native speakers' swing.

At the time of writing, I cannot account for this difference between the performances of native speakers and non-native speakers between the two studies. It may well be that my comments in Section 4.1 above concerning the greater sensitivity of native speakers to the current settings for normal were premature: the evidence from Study 2 Extract 2 seems to suggest that non-native speakers are in certain circumstances more sensitive than native speakers to the in-context/in-isolation hearings adopted in these studies.

A strong theme that emerges from both studies is that contextual factors – whether they be situational or prosodic (the current settings for normal) – can neutralise the perceived meaning potential of the prosody of an isolated extract. In this case, contextual conditions were such that in-isolation prosodic choices that seemed to cue *anger*, did not do so in-context. In Study 1 the power of hearing the wh-questions in-context was great enough to reverse previous judgments of *anger*. In Study 2 the power of hearing the wh-questions in-isolation was great enough to reverse previous judgments of *neutral*. These reversals of judgment were achieved without kinesic information: the aural evidence – including greater knowledge of the current settings for normal in this specific context – together with some situational information – sufficed to produce these reversals.

6. ATTITUDE TABLES REVISITED

Attitude tables (such as Tables 1-3 above) even with contextual factors included in them, cannot adequately account for attitudinal meanings in spontaneous speech, because they do not (and cannot ever) contain the required information on what the current prosodic settings for 'normal' are in any particular context for a particular speaker. To account for how attitudinal meanings are cued and perceived, one has to start with contextualised language, and with a full characterisation of those factors that influence speakers' behaviour and hearers'

perceptions of that behaviour. These factors are not amenable to being placed in cause-and-effect relationships in an Attitude Table.

5. REFERENCES

1. Brazil, D. (1994) *Pronunciation for Advanced Learners of English*. Cambridge: Cambridge University Press.
2. Cauldwell, R.T. (1997) Tones, Attitudinal Meanings, and Context. *Speak Out! Newsletter of the IATEFL Pronunciation Special Interest Group* 21: 30-35
3. Crystal, D. (1969) *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.
4. Crystal, D. (1975) *The English Tone of Voice*. London: Edward Arnold.
5. Crystal, D. (1995) *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University Press.
6. Halliday, M.A.K. (1994) *An Introduction to Functional Grammar*. [2nd Edition] London: Edward Arnold.