This is a repository copy of *Phonetics and prosody in conversation*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/88518/

Version: Accepted Version

Book Section:
Walker, G. (2012) Phonetics and prosody in conversation. In: Sidnell, J. and Stivers, T., (eds.) The Handbook of Conversation Analysis. Blackwell Handbooks in Linguistics . John Wiley and Sons, Ltd , Chichester, West Sussex , pp. 455-474. ISBN 9781444332087

https://doi.org/10.1002/9781118325001

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

This is the author's final version of the chapter available via

http://dx.doi.org/10.1002/9781118325001.ch22

Some small changes may have occurred after this version was sent to publication. The final published version should be consulted before quoting or discussing in detail.

Phonetics and prosody in conversation*

Gareth Walker

University of Sheffield

1 Introduction

There is a sense in which anyone engaged in the analysis of audio recordings of talk-in-interaction is analysing its phonetic design: there is, after all, no talk-in-interaction without phonetic design. The elaborate system of notation developed principally by Gail Jefferson for the transcription of unscripted talk-in-interaction provides evidence of sociologists recognizing the potential interactional relevance of phonetic details. That the Jefferson system is still so widely used shows that conversation analysts – whether trained in phonetics or not – routinely register the potential interactional import of phonetic detail to the conduct of social, spoken interaction. Since Goldberg (1978) a body of work has built up in which techniques of auditory and acoustic phonetics are applied to the study of talk-in-interaction in ways which aim to be

* Thanks to the editors and Traci Walker for comments on earlier drafts of this article; their comments raised many interesting issues, only some of which I have been able to address here.

consistent with the principles of CA, identifying interactional relevancies of the features described. It is some of that work, and some issues which surround it, which is described here. Analytic principles of that work are described in section 2; some outcomes of analyses following those principles are outlined in section 3; issues in representing talk-in-interaction on paper are discussed in section 4; some future directions and challenges are described in section 5.

2 Analytic principles

Technical phonetic analyses of conversation of the type described in this chapter (see also e.g. Couper-Kuhlen & Ford, 2004; Couper-Kuhlen & Selting, 1996) has certain characteristics. Two strands of phonetic analysis can be identified: *auditory* (where analysis is conducted by means of careful, repeated listening) and *acoustic* (where machine-generated acoustic records are analysed). Auditory analysis is more common, but a combination of auditory and acoustic analysis is increasingly widespread. Where both forms of analysis are carried out, these are done reflexively, i.e. through repeated comparison of what can be heard with what can be identified in acoustic records. This section describes these two forms of phonetic analysis. A detailed consideration of what is involved in auditory and acoustic phonetic analysis is beyond the scope of this chapter. The purpose of this section is to explain some basic principles of phonetic analysis as relevant to the study of conversation that will assist a reader new to this mode of enquiry in following later sections of this chapter, in understanding the literature of the field and in their own analyses.

2.1 Auditory analysis

Auditory analysis involves careful, repeated listening to audio recordings. Claims are based on what can be heard in the audio/video recordings and should be underpinned by at least a basic understanding of the speech production process. Listening is done parametrically: “[p]honetic observation begins by listening to speech in terms of independently varying auditory and movement parameters and not in terms of

unanalysed, static postures and transitional glides" (Kelly & Local, 1989a: 30). This involves being able to make the analytic separation between one vocal organ (larynx, tongue, lips, velum, etc.), or a part of that vocal organ, and all others. For instance, it requires separating out the ways in which air is being modified as it passes through the larynx to produce e.g. voiced or voiceless phonation, creaky voice, breathy voice etc., from the precise positioning and movements of the vocal organs in generating different sound types, e.g. plosives, fricatives, approximants, resonants, etc. with different specific qualities. Parametric listening also involves listening for features of pitch (e.g. relative height, as well as the direction and magnitude of any pitch movement), loudness, and duration (and its derivatives: articulation rate, rhythm and tempo), among others. In practical terms parametric listening usually involves the analyst combining careful listening with kinaesthetic, empathetic awareness in an attempt to replicate (often silently) a speaker's vocal gestures: a process that Kelly & Local describe in detail (1989a). Once the analyst has made any necessary adjustments to the arrangement or movements of the vocal organs, conclusions can be attempted concerning how the original utterance was produced.

Parametric listening is an especially useful analytic technique for the study of conversational materials for two main reasons: first, it cannot be known in advance which details will be of importance and second, a wide range of phonetic details have been shown to be relevant to the organisation of interaction.[1] Parametric listening sits comfortably alongside CA since both are less theoretically guided approaches to spoken data and favour the 'unmotivated looking' for order over the explicit testing of ready-formed hypotheses or theoretical models. Open-minded, parametric-impressionistic listening has yielded considerable insights into the role of fine-grained phonetic details which would have otherwise gone unnoticed. Some of these details are described in section 3.

[1] These are also reasons why in most cases in this chapter the term *phonetic* is used rather than *prosodic*. The terms prosody/prosodic are problematic in certain respects: their meaning is vague, and they are not used consistently in the literature (see Peppé 2009). Their use in general discussion suggests that some features, e.g. of pitch, loudness, speech rate, etc., can be legitimately investigated without engaging in detailed consideration of other aspects of the speech signal i.e. articulation and phonation. While much work shows that features typically dubbed prosodic are important, a considerable amount of research shows that other, more general phonetic features are relevant.

2.2 Acoustic analysis

Analysts working on the phonetic details of conversation refer increasingly to acoustic measures and records, often to support and supplement auditory judgments. The most commonly used types of acoustic records in studying the phonetics of conversation are outlined here (further detail can be found in e.g. Harrington, 2010; Ladefoged, 2003).

Measures and traces of fundamental frequency (F_0) have become commonplace in the literature. F_0 corresponds to the rate of vocal fold vibration and is the acoustic correlate of pitch. F_0 is typically expressed in Hertz (Hz): the higher the Hertz value, the higher the F_0 and, roughly speaking, the higher the perceived pitch. Figure 1 shows an F_0 trace of an utterance, with word-labels. As for all acoustic records in this chapter, time is shown on the x-axis; frequency is shown on the y-axis.[3] A practical issue in the interpretation and presentation of acoustic information is the extent to which it reflects participants' likely percepts. Placement of talk within a speaker's range, rather than a raw measure of F_0, has been shown to have interactional significance (Couper-Kuhlen, 1996; Local, 2004). A better visual representation of F_0 is therefore one which is scaled to the speaker's baseline and topline pitch values, as in Figure 1. In this case the speaker's baseline and topline pitch values were estimated on the basis of maximum and minimum F_0 values in a one minute sample of conversational speech produced by this speaker; the median F_0 value in such a sample can be used as an estimate of the middle of a speaker's range. Figure 1 uses a logarithmic frequency scale to take into account the non-linear perception of pitch (listeners perceive a greater change in perceived pitch at lower frequencies). Comparative F_0 measures are usefully presented using a non-linear scale such as semitones (ST).[4]

[3] Sound files are currently available via http://gareth-walker.staff.shef.ac.uk.
[4] An increase of 12 ST (1 octave) represents a doubling in frequency.

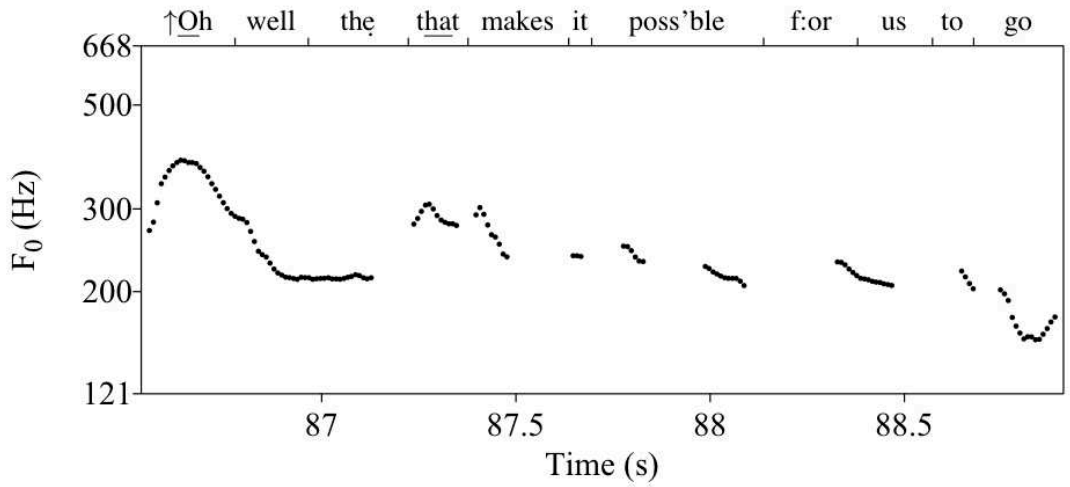


Figure 1: F_0 trace of an utterance by a British female produced in the course of an unscripted telephone call (Holt corpus)

Duration is most readily and accurately measured in the *sound pressure waveform* (or more usually simply *waveform*), as shown in the lower part of Figure 2; changes in pressure are represented by vertical displacement. As well as raw duration measures of phonetic events e.g. periods of silence (Stivers et al., 2009) and intervals between isochronous 'beats' (Auer, Couper-Kuhlen, & Müller, 1999), it is also possible to provide quantified measures of articulation rate. Many phonetic events have characteristic visual properties on a waveform some of which are identified in Figure 2. A sample spectrogram is shown in the upper part of Figure 2, time-aligned with the waveform.[5] Spectrograms show frequency on the y-axis and intensity as relative darkness. Spectrograms are especially useful for the inspection of articulatory and phonatory characteristics, and how these features change over time. Spectrograms are used to provide corroborative acoustic evidence for articulatory and phonatory characteristics of interactional relevance in articles including Local (2004); Local and Walker (2004); Ogden (2001, 2004); Walker (2004).

[5] Since the waveform and spectrogram are time-aligned, some features are labelled only in one record; by looking directly above/below the labelled portion it should be possible to identify the relevant features in each record.

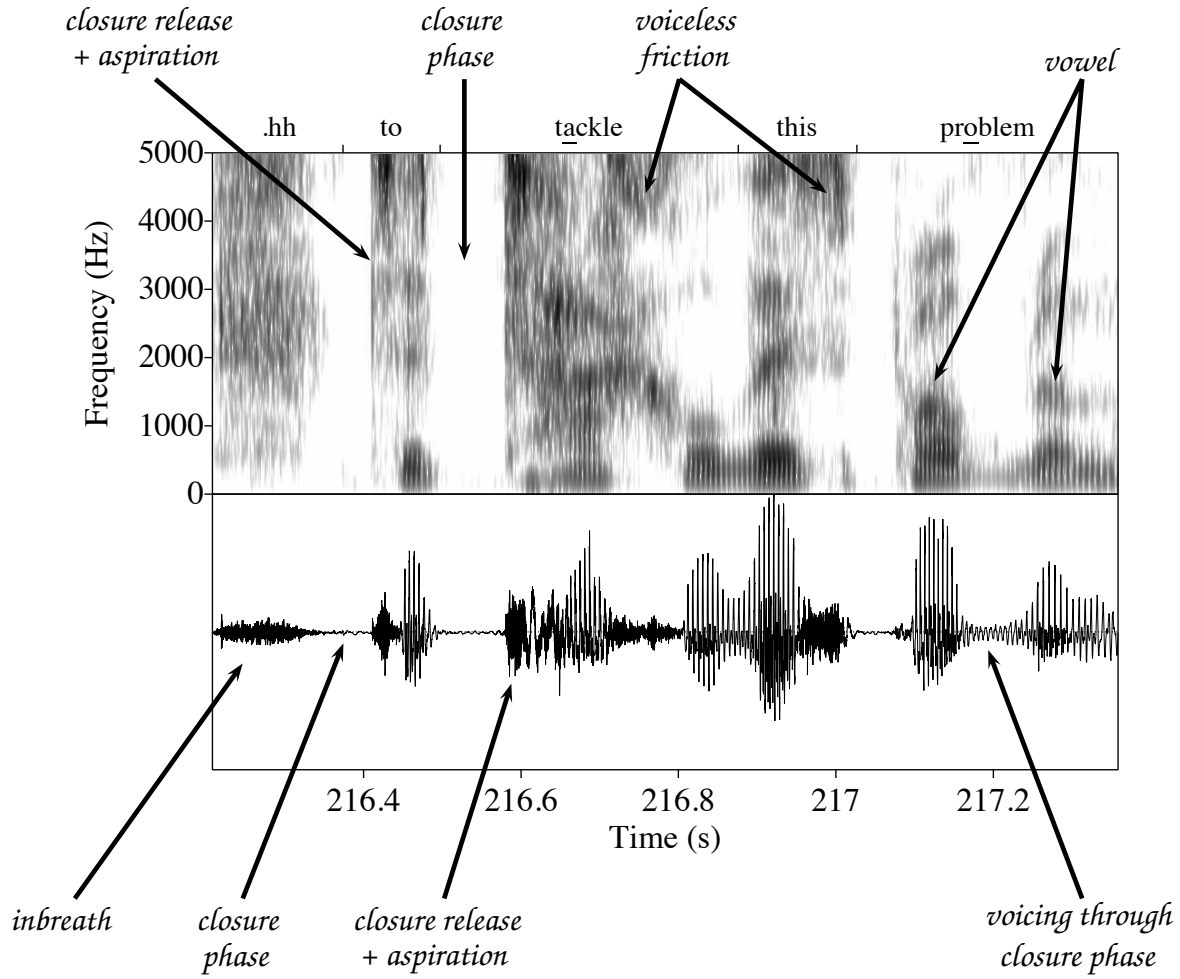


Figure 2: Waveform (lower part) and spectrogram (upper part) of an utterance by a British female produced in the course of an unscripted radio interview (*Today* programme)

It should be noted that phonetic analysis of naturally occurring materials is usually conducted in the face of several confounding factors. The presence of overlapping talk, background noise, recordings of limited overall quality, and speakers at different distances from microphones are among the challenges which have to be faced when analysing naturally-occurring data. Typically these challenges are all the greater when computer-based acoustic analysis is being performed. Some more localised features may prove problematic e.g. changes in phonation type leading to unreliable measures of fundamental frequency, and different segmental qualities affecting intensity measures. More often than not, meeting these challenges involves conducting only those analyses where there are good reasons for believing that they will produce robust, meaningful measures.

It is not a requirement that acoustic analysis be carried out. Numerous important studies setting out relationships between phonetic design and the organisation of talk-in-interaction (some of which are referred to in this chapter) have no reported acoustic component. However, when used appropriately, acoustic analysis can provide important objective corroborative evidence for what can be subjectively perceived in the speech signal, as well as providing the basis for visible representations of relevant features. Potential limitations of acoustic analysis (including, for instance, the lack of agreed techniques of measurement/ representation for certain aspects, and possibly complex relationships between measurements/representations and auditory percepts) make continued reference to *what can be heard* essential. Indeed, in certain analytic contexts auditory analysis alone may be appropriate.

2.3 Applying auditory and acoustic phonetic analysis

Research in phonetics has given us a rich set of resources with which to provide detailed, accurate descriptions of the pronunciational features of utterances. The methodology outlined in the preceding section requires a systematic analysis of the phonetic properties of utterances: systematic in terms of the details considered, and their mode of description. As an example of the kinds of analyses which can be brought to bear on conversational data consider Fragment 1, and particularly the end of Curt's "out there" in line 17.[6]

(1) Schegloff (1987: 103)

```
1  Curt:    (W'll) how wz the races las'night.
2           (0.8)
3  Curt:    Who w'n [th'feature.]
4  Mike:             [Al won,    ]
5           (0.3)
6  Curt:    [(who)]=
7  Mike:    [Al.  ]=
8  Curt:    =Al did?
9           (0.8)
10 Curt:    Dz he go out there pretty regular?
```

[6] The corresponding sound clip can be accessed via http://www.sscnet.ucla.edu/soc/faculty/schegloff/sound-clips.html.

```
11          (1.5)
12 Mike:    Generally evry Saturdee.
13          (1.2)
14 Phyllis: He wins js about every Saturday too:.
15 Ryan:    Bo[: Bo!
16 Curt:      [He- He's about the only regular<he's about
17          the only good regular out there.'z Keegan still go
18          out?=
19  Mike:   =Keegan's, (0.2) out there (,) he's, He run,
20          (0.5)
21  Mike:   E:[r he's uh::     ]
22  Gary:     [Wuhyih mean my:,]
23  Gary:   My [brother in law's out there,]
24  Mike:      [doin real good this year'n ] M'Gilton's
25          doin real good thi[s year,
26  Curt:                     [M'Gilton still there?
```

There is audibly rising pitch on "out", and low-level pitch on "there" near the bottom of Curt's normal speaking range: see Figure 3 which shows a combined F_0 and intensity trace of part of Curt's turn. In terms of articulation rate, "there" is not especially quick. This auditory impression is borne out by comparing its duration with that of Curt's earlier "feature" (line 3). "Feature" and "out there" both have similar (though not identical) segmental and syllabic make-up and both carry the major accent ('pitch peak') of their turn-constructional unit on the first syllable (Schegloff, 1987: 106). On the basis of conservative segmentation, "feature" measures 322 ms, while "out there" measures 373 ms. It can be heard that "there" has *diminuendo* (decreasing) loudness characteristics: it gets audibly quieter over time. As can be seen in Figure 3, the drop-off in intensity over "there" is 7.5 dB; a drop in sound-level of 10 dB would be roughly equivalent to a halving of loudness (Laver, 1994: 502). Diminuendo is regarded as one phonetic end-of-turn marker (Duncan, 1972; Nolan, 2006); preliminary research shows that points of possible completion which are not (to be) treated as complete do not exhibit diminuendo (Local & Walker, 2010).[7]

[7] Note the loudness characteristics of Curt's production of "regular" and into the next word: the drop-off in intensity here is only 2.7 dB.

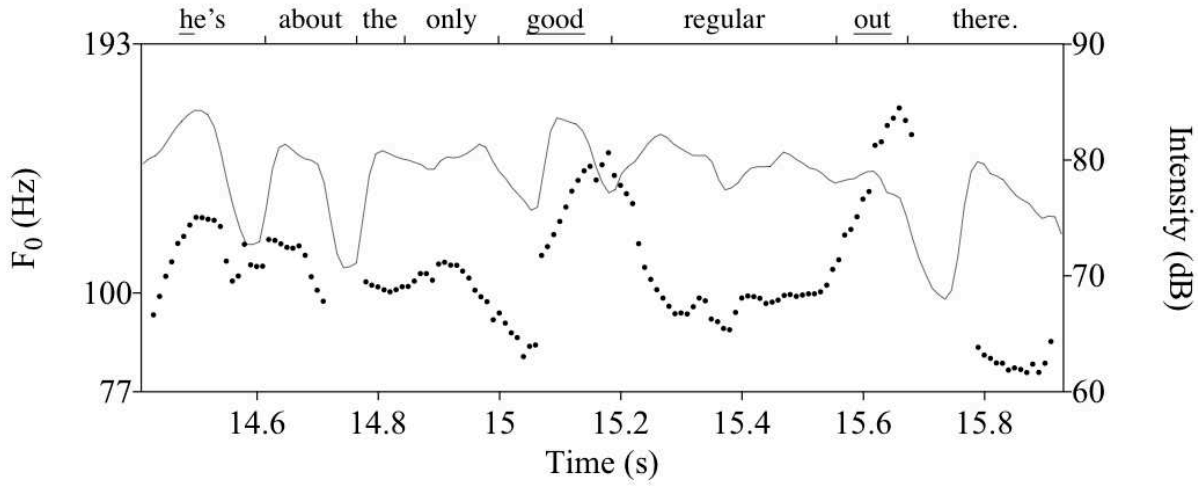


Figure 3: Combined F_0 trace (dotted line, left axis) and intensity trace (solid line, right axis) of part of Curt's turn in Fragment 1, lines 16-17

This is not the first time the phonetic characteristics of this part of Curt's turn have been described. Curt's "out there" has been described previously as an example of a 'rush-through' (Schegloff, 1987). The rush-through is an example of a phenomenon which comes into being through its phonetic constitution, but which has been left wanting in terms of an account of its precise phonetic characteristics (Schegloff, 2005). In a general description Schegloff describes the 'rush-through' as

> a practice in which a speaker, approaching a possible completion of a turn-constructional unit, speeds up the pace of the talk, withholds a dropping pitch or the intake of breath, and phrases the talk to bridge what would otherwise be the juncture at the end of a unit. . . Here the turn-extension device is initiated near the otherwise-possible-end of the turn, rather than at its beginning. (1982: 76).

Schegloff describes part of Fragment 1, line 17 as follows:

> Curt methodically organises the production of his talk, that is, the first component of his turn, to provide for the addition of another component. Using a device we can call the "rush-through" (Schegloff, 1982), he speeds up the talk just before possible completion of the first turn-unit ("there" does not have the "drawl" or sound stretch often found in last words or syllables); he omits the slight gap of silence which commonly intervenes between one unit and another, reduces the first sound of what follows to its last sound ("z") and thereby "rushes" into a next turn-constructional unit, interdicting (so to speak) the otherwise possibly relevant starting up of talk by another at that point. (Schegloff, 1987: 104).

While there is reduction of "does", with no initial closure or vocalic portion, and no silence

between “there” and “does” the observations concerning duration (speeding up) and pitch (non-low pitch) are wayward, and certain other relevant phonetic characteristics (e.g. concerning loudness) are absent from the description.

The mismatch between what can be heard (and measured) in the audio and Schegloff’s “quasi-phonetic” descriptions (Schegloff, 2005:450) is especially problematic for anyone attempting to locate and understand ‘rush-throughs’ in their own materials. Anyone trying to do that would have to make a choice between following Schegloff’s description, or following the details of his exemplar. A more robust mode of analysis and/or description would yield, as one of its outcomes, an account of the relevant phonetic properties which could be more readily applied by others to their own materials (for an attempt at such an account see Walker, 2010). A more thorough-going account of the phonetic details of Curt’s turn may have an impact on how we understand what Curt is doing in his turn. Schegloff’s observations notwithstanding, any manipulation of the transition space (see Clayman, this volume) is going on after the end of “there”, which has the phonetic characteristics of a routine turn-ending (on low final pitch as one kind of turn-terminal pitch configuration, see Ford & Thompson, 1996, Local, Kelly & Wells, 1986; on unit-final slowing down, see Local et al., 1986, Turk & Shattuck-Hufnagel, 2007; on diminuendo see Local et al., 1986, Obeng, 1991). Curt seems to be designing the start of his next unit to come off as something like a retrospective early start: an action which may be related to, but distinct from, the prospective speeding up Schegloff’s nomenclature suggests.

3 Outcomes

This section draws together – and in some cases builds on – some of the findings from studies conducted over the last 30 years or so which investigate the relationship between the phonetic design of talk and the organisation of interaction. It is important to bear in mind that these discoveries have been made through combining parametric phonetic analysis and CA, and that neither approach on its own could have yielded these findings, or set them out in such detail. Phonetics provides CA

with the analytic and descriptive techniques required for the careful and systematic description and analysis of its main source of materials: spoken interaction. CA provides phoneticians with the means to describe and analyse the sequential organisation of talk, and a framework for establishing the interactional relevance of details of speech production. A consideration of the potential or actual interactional relevance of phonetic details is almost entirely absent from other branches of phonetics. CA's emphasis on the analysis of naturally-occurring talk-in-interaction is also relevant. Speech elicited in a laboratory remains one of the main sources of data in phonetics, although 'lab speech' – even when it is not scripted – does not generally require subjects to handle the same range of interactional tasks as naturally occurring talk-on-interaction; participants' management of those tasks therefore cannot be analysed on the basis of those materials. Three tasks are discussed in this section: the management of turn-taking, the marking of relationships within turns, and the marking of relationships between turn-constructional units (TCUs).

3.1 Turn-taking

Managing entry to and exit from talk is a key task for participants engaged in interaction: when speaking, they must signal to co-participants points of possible turn transfer and when listening they must monitor for such points in order to begin talk appropriately. Since the publishing of Sacks, Schegloff, and Jefferson (1974) a body of work has built up describing the role of phonetic detail in the management of turn-taking: see e.g. Auer et al. (1999); Ford and Thompson (1996); Fox (2001); Local et al. (1986); Local, Wells, and Sebba (1985); Selting (2000); Wells and Corrin (2004); Wells and Macfarlane (1998); Wells and Peppé (1996). Although there is variation in the precise analytic focus of these studies – and the relative importance ascribed to the phonetic design of talk in terms of its role in managing turn-taking – they routinely find that phonetic characteristics of the talk contribute to the status of an utterance as transition relevant, or not. Features described include prosodic (pitch, loudness, duration and rhythm) and ‘non-prosodic’ (articulatory, phonatory) features. Ogden (2001, 2004) shows how creak phonation and glottal stops figure in the management of turn-taking in Finnish. Ogden argues that creak is associated with turn-yielding, and glottal stops with turn-holding: while creak can mark talk out as transition relevant, a glottal stop at

the end of an otherwise possibly complete TCU marks that there is more to come from that speaker (cf. Local & Kelly, 1986, Local & Walker, 2005).

One apparently overlooked area where phonetic analysis could contribute is in understanding the role of inbreaths in the management of turn-taking (but see Schegloff, 1996, 2000 on sequential aspects of pre-talk inbreaths). Inbreaths can be produced in such a way as to work against a co-participant starting up their own talk by being produced as close as possible to the end of that participant's preceding TCU and the beginning of the next, without initial or final glottal closure, and with high energy; an example is shown in line 1 of Fragment 2.

(2) Holt.SO.88.II.1.3, 8:23

```
1 Hal: -->  But um (.) No. it eh it's marv'lous realli-.hh YOU DID'↑n:t
2           e-Did Marian: (.) ever ask you- it's too late now b't
3           d'Marian as'you 'bout that Scout da:nce?
4                 (0.3)
5 Leslie:   N-uh no:?
```

Hal brings his talk in line 1 towards a point of possible completion. There is audible evidence of articulatory reduction and temporal compression of his "realli-" (line 1) of a similar order to the reduction and compression observed in 'rush-throughs' (Walker, 2010).[8] The inbreath begins in maximally close temporal proximity to the preceding "realli-", and the following talk is produced in very close proximity to the inbreath: see Figure 4.[9] That Hal's inbreath has distinctive loudness characteristics is supported by comparative intensity measures. The peak intensity for the inbreath in Fragment 2 is 65.5 dB. This is high relative to whereas of all the other inbreaths from Hal in this call (n = 10). The mean of these peak intensity values is rather lower at 63.6 dB. Moreover, only one has a higher peak intensity (and only slightly higher: 66 dB). The temporal proximity of the inbreath to the preceding talk and of the following talk to the inbreath minimises the time in which Hal is not 'making noise', and therefore in which his co-participant can begin her own talk 'in the clear'. As well as making the inbreath more readily perceptible to his co-participant (and therefore potentially increasing its

[8] The '-' symbol in the transcription notwithstanding there is no audible or acoustic evidence of a final 'cut-off'.

[9] A certain amount of delay between (i) the offset of voicing and the inbreath, and (ii) the end of the inbreath and the onset of voicing is inevitable due to the switch from egressive to ingressive airflow, and back again.

interactional relevance), the quality of the inbreath is "characteristic of some disjunctive next move" (Drew & Holt, 1998: 507-8; see also Jefferson, 1981: 72-75 on inbreath-initial topic shifts). In this case Hal's next move is to produce talk on a different topic.[10] The inbreath therefore seems to work alongside, and potentially in system with, other lexico-grammatic markers of topic-shift e.g. "anyway…", "getting back to X…", and so on. The speeding up at the end of the preceding word allows Hal to get into this talk-projective inbreath (if not the projected talk which follows it) sooner than his co-participant might have anticipated (cf. 'rush-throughs').

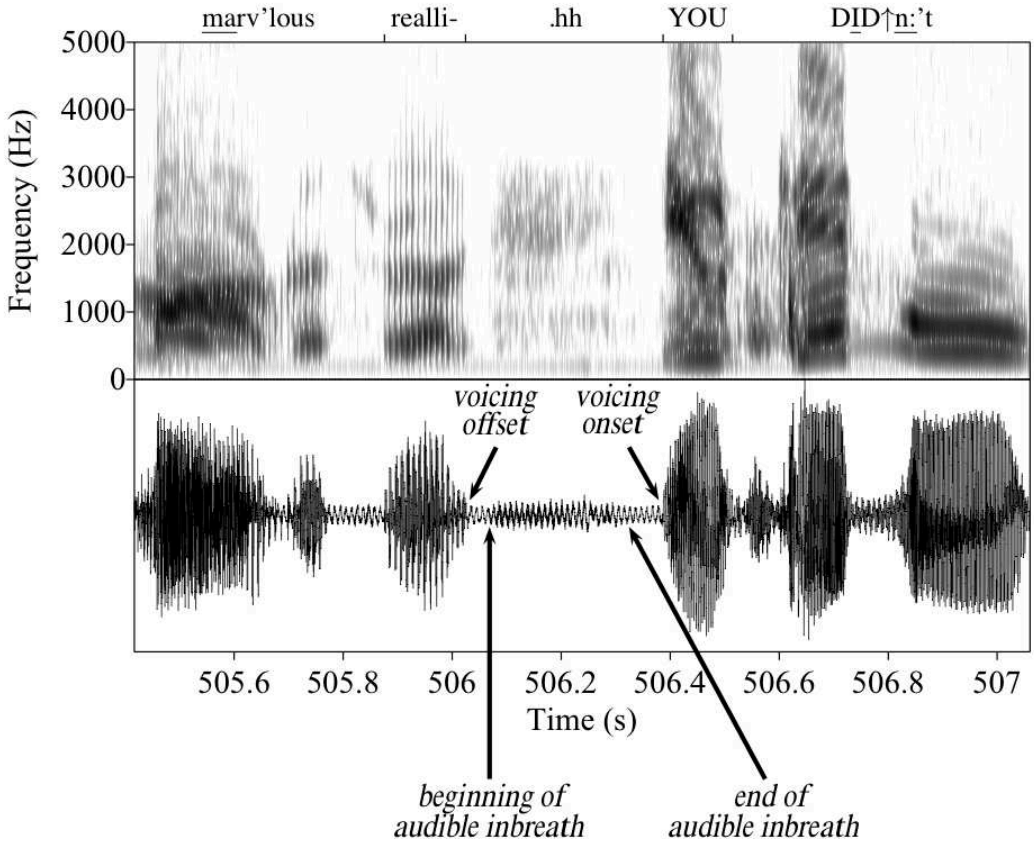


Figure 4: Labelled waveform (lower part) and spectrogram (upper part) of Fragment 2, line 1; background noise, rather than vocal behaviour, is responsible for the quasi-periodic wave with a F_0 of approximately 95 Hz

3.2 Relationships within turns

[10] Note also the occurrence of self-repair in Hal's post-inbreath talk; on the regular occurrence of self-repair in topic-initial turns, see Schegloff (1979).

The phonetic design of many turns at talk, and of their component parts, is such that they are hearable as cohesive chunks. In Fragment 1/Figure 3, Curt's "he's about the only good regular out there" (lines 16-17) is produced as a single cohesive chunk: there are no sudden, disjunctive changes in pitch, loudness, articulation rate, or voice quality, and no glottal or supraglottal 'cut-offs' (see also Walker, 2007). Such disjunctive features can display that some different line is being taken from that which the developing talk projects (Couper-Kuhlen, 2004b; Goldberg, 1978, 2004; Jasperson, 2002; Local & Walker, 2004). In certain sequential environments some of these features can mark out that a previous line is being resumed. Consider Fragment 3, taken from a broadcast radio interview between interviewer James Naughtie (JN) and Lord Falconer (Fal).

(3) Today 5 April, 2004, 6:20

```
1 Fal:      it's positive and it's got to be (.) positive we should not
2           .h be ashamed of saying it
3                 (.)
4 JN:       .mthhh (0.8) do we need just to pick up on another of: Mister
5           Duncan's questions befo- before I move on .hhh (.) do we
6           need to have a policy in this country which says that all
7           those who are .hhh uh-m likely to be persecuted in their
8           home countries (.) must be given refuge here
```

In line 4 JN makes a start on his next question with "do we need", before breaking off and producing talk on a different line. His "do we need" is recycled in lines 5-6 and his question brought to completion. Embedded within the bounds of JN's TCU is talk (marked with dashed underlining) which is *parenthetical* to his question. It is, in large part, the phonetic design of JN's turn which marks out this insertion as parenthetical to the main question (see also Local, 1992; Mazeland, 2007). The phonetic design of the talk handles three main task in this regard. First, phonetic design marks out where the parenthetical talk begins. One change is that JN's parenthetical talk is noticeably faster than the talk which precedes it or which follows it. (His turn-initial "do we need" is produced at a rate of 5.4 syllables per second [syll/s], his parenthetical talk is produced at a faster rate of 6.2 syll/s, and his post-parenthetical talk up to the end of his turn is produced at a rate of 5.1 syll/s.)

That JN is in the midst of a parenthetical is signalled by the production of the parenthetical talk low in his pitch range (pre-parenthetical mean F_0 = 160 Hz, max. = 206 Hz, min. = 111 Hz; parenthetical mean F_0 = 98 Hz, max. = 136 Hz, min. = 72 Hz; post-

parenthetical mean F_0 = 106 Hz, max. = 231 Hz, min. = 231 Hz). Second, phonetic design marks out where the parenthetical talk ends. The “on” of JN’s “before I move on” (line 5) is produced with falling-rising pitch. Local (1992: 278) reports that what he labels *self-interrupting talk* – of which that in Fragment 3 is one kind – is routinely characterised by rising pitch at its end. Like Hal’s inbreath in line 1 of Fragment 2, JN’s inbreath immediately after “on” projects the production of talk on some different line from that immediately prior. The final rising pitch and a high energy inbreath, together with the arrival at a point of possible syntactic completion for the parenthetical project most strongly that the parenthetical has ended and the line taken earlier is about to be resumed. Third, the phonetic design marks out the post-parenthetical talk as a return to the talk begun earlier. JN’s second, resumptive “do we need” is audibly louder, slower and higher in pitch than the parenthetical talk which it follows (for the resumptive “do we need” mean syllable peak intensity = 82.1 dB, mean F_0 = 173 Hz, articulation rate = 4.5 syll/s; for the preceding “before I move on” mean syllable peak intensity = 74.7 dB, mean F_0 = 93 Hz, articulation rate = 6.1 syll/s). Importantly, the loudness, articulation rate and pitch characteristics also mark out this resumptive “do we need” as picking up the earlier line started at the beginning of JN’s turn then postponed (turn-initial “do we need” mean syllable peak intensity = 81.4 dB, mean F_0 = 160 Hz, articulation rate = 5.4 syll/s).

The point at which the parenthetical begins in Fragment 3 is not simply marked out by a break in the unfolding grammatical structure of the turn-in-progress. The lexical choices JN makes do not clearly identify the parenthetical as a parenthetical: he could be asking Lord Falconer the question “do we need just to pick up on another of Mr Duncan’s questions before I move on?” This is not the case and Lord Falconer does not treat JN’s talk as such a question. However, this only becomes clear if phonetic design is considered as an integral part of the unfolding turn.

3.3 Relationships between turn-constructional units

Phonetic design can mark out how the current turn or TCU relates to preceding talk: phonetic design establishes certain sorts of syntagmatic relationships between turns and

parts of turns. On the basis of question-answer and telephone-call closing sequences Goldberg (1978, 2004) argues that relationships between utterances can be expressed through shifts in amplitude (and, by extension, in perceived loudness): a downward shift (decrease) in the amplitude of successive utterances *affiliates* the current utterance to a prior, whereas an upward shift (increase) *disaffiliates*. Sequence initiation, she argued, is accompanied by raised amplitude. Couper-Kuhlen (2004b) argues that increases in pitch and loudness are among the prosodic features which can mark out the current turn not as a continuation of the action which went before, but as the beginning of a new course of action. Local (2004) argues that a particular, stable set of phonetic design features attendant on “and-uh(m)” marks out what follows as a resumption of something earlier in the interaction (see Bolden, 2009 for a review of the literature on resources for managing such connections in a range of languages). This, Local argues, is in contrast to the sorts of variability observed in the phonetic design of turn-initial “and” where talk is being built incrementally on the immediately prior talk (Heritage & Sorjonen, 1994). Local argues that co-participants orient to the turns beginning with “and-uh(m)” as returns to a prior activity, principally through their alignment with and engagement in, the activity being resumed. Audible design features of the ‘resuming’ “and-uh(m)s” Local describes include: the production of “and-uh(m)” with initial creaky voice and/or glottal stop, a full (non-reduced) vowel in “and”, roughly level pitch in the middle of the speaker’s pitch range, a relatively slow rate, and no break between the two syllables. Although not discussed by Local, “but-uh(m)” seems to play a similar role when accompanied by equivalent phonetic features. An example is shown in line 15 of Fragment 4; Skip and Fred are work colleagues. The fragment immediately follows the identification/recognition sequence at the start of the call.

(4) Holt.U88.1.10, 0:08

```
1  Ski:    Did you go back to wo:rk,h
2                 (0.2)
3  Fre:    I've gotta me:ssage to ring Raymond Smi[th.
4  Ski:                                          [Oh: yes:.
5  Ski:    That's alright I just wanted to make sure: (.) whether
6          you'd p'hh gone back or no[t.h
7  Fre:                              [Yes I did. No[I got that=
8  Ski:                                           [.hhhhhhh.p
9  Fre:    =thanks 'n I, I've also heard about th'of course about
10         the cash ↓in toda:[y.↓
11 Ski:                      [gYes::.
           ((c. 20 lines of transcription omitted in which recent payments made are discussed))
```

```
12 Ski:      =di:d..hhh But anyway n-we've still got s'more t'go=
13 Ski:      =ou:t,h .hh-h[h
14 Fre:                  [Ye[s.
15 Ski: -->                 [But uh dHave you rung Raymond,[h-h-h]=
16                                                         [(0.4)]=
17 Fre:      No
```

“But uh” (line 15) prefaces talk which connects not to the immediately prior talk but to something earlier which did not run its course (cf. Local, 2004). Skip’s enquiry at line 1 is one step in a move towards his offer of the reason for his call (see Schegloff, 1986 on the organization of call openings). This move is derailed somewhat by Fred’s response in line 3. After talk on other matters, Skip returns to the issue of whether Fred has called Raymond Smith in his “But uh”-prefaced turn in line 15 with the reason for the call to which his enquiry at line 1 was conceivably heading: whether or not Fred had called Raymond Smith. (It turns out that Raymond Smith is “rather anxious”, as Skip puts it, that Fred calls him either that evening or “first thing tomorrow”; see also Schegloff, 2010 on “uh(m)” as a preface to reasons for calling, including those which, as in this case, are delayed until some way into the call.) That such a back-connection is underway is signalled by the characteristic phonetic design of the “But uh”. Figure 5 provides visual representations of this part of Skip’s turn. Most notably, and relevant to this back-connecting work, Skip’s “But uh” exhibits (i) a long period of aspiration on release of the final closure of “but” (at c. 46.26s), followed by voicing for the following vowel (at c. 46.33s) without any intervening silence, (ii) ‘equal-equal’ rhythm: both syllables have equal prominence, (iii) a pitch step-up from the prior talk: Skip’s prior “go out” has a final F_0 of 151 Hz and “but” begins at 226 Hz (a difference of 7 ST), (iv) broadly level pitch (range = 1.93 ST) near the middle of his range and 12.1 ST above his baseline pitch. The case for the phonetic design of Skip’s “But uh” handling back-connecting work is supported by the equivalence of the phonetic features here to those described by Local (2004), and also by the difference in the phonetic design of this “But uh” (and other, similar tokens) from other cases of “but uh(m)” described in the literature. Schegloff (2009) describes the role of “but uh(m)” (and other, similar constructions) in (re-)exiting a sequence. The exemplar of sequence-exiting “but uh(m)” Schegloff provides (Schegloff, 2009: 370-1) is interestingly different in its phonetic design from the case presented here, particularly in its audible pitch characteristics, and more particularly still, the placement of this utterance in the speaker’s pitch range. Like Skip’s “But uh”

in Fragment 4 above, in Schegloff's sequence-exiting example "But u:m" is produced with broadly level pitch (range = 1.3 ST). However, the sequence-exiting example is hearable as much closer to the bottom of the speaker's range in that case (3.4 ST above his baseline pitch). It is perhaps unsurprising that the two tokens should differ in this manner, since the use of raised pitch at the start of a new sequence has been documented elsewhere (see e.g. Couper-Kuhlen, 2004). One potential interactional relevancy of different phonetic designs for back-connecting and sequence-exiting "but uh(m)" is that back-connnecting and sequence-exiting "but uh(m)"s occur in equivalent sequential environments, i.e. the potential end of one sequence and start of another: there is nothing intrinsic to "but uh(m)" to indicate who might start that next sequence, or how that sequence might relate to any prior talk. However, the phonetic design of the token *can* indicate such things, as in these cases.

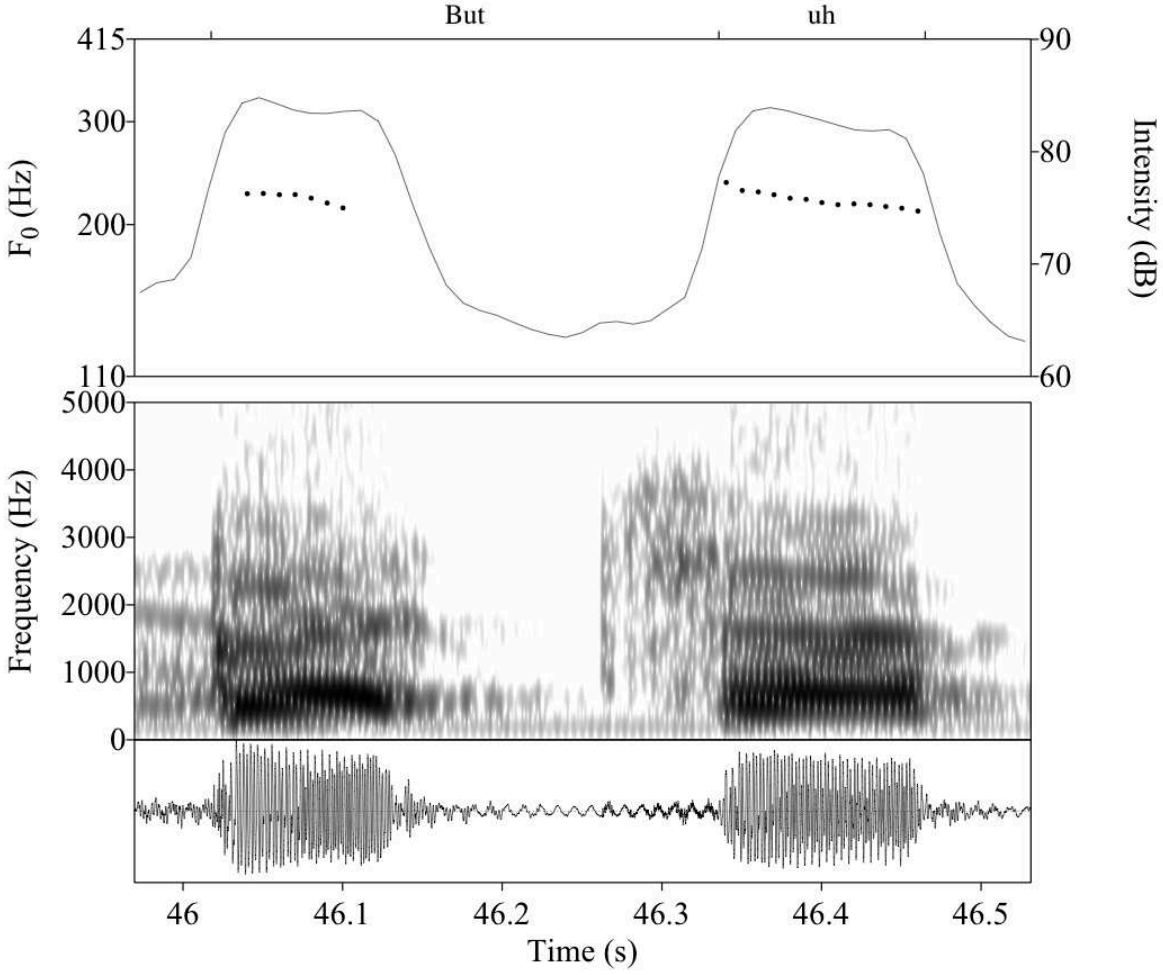


Figure 5: Waveform (lowest part), spectrogram (above), and (topmost part) combined F_0 (dotted line, left axis) and intensity trace (solid line, right axis) for part of Fragment 4, line 15

4 Transcription

A new, or even modified system of notation has not been a central goal of work on the phonetics of conversation (but see Wells & Local, 2009 for an attempt to capture in a transcription the outcome of analyses of turn projection). However, since transcribing constitutes a central activity in both CA and phonetics it is relevant to discuss at least some issues here. The notation system discussed here is that developed principally by Gail Jefferson since it is the most widely used system in CA (see Hepburn and Bolden, this volume for an explanation of the main conventions).[11]

The Jefferson system has a number of redeeming features.[12] It has been applied to a range of types of non-elicited data, including (but not limited to) telephone interactions, face-to-face interactions, everyday conversation, institutional talk of various kinds, two-party interactions, multi-party interactions, audio-only data, and video data (with appropriate extensions for visual features). The transcriptions are consistent with regard to the representation of certain details, and especially details of sequential organisation (e.g. periods of silence and simultaneous talk). The notation system provides for the capturing of certain finer phonetic details by way of symbols, e.g. upward and downward pointing arrows to represent pitch movements to "especially high or low" pitch (Jefferson, 2002: 1379), various punctuation marks to indicate aspects of intonation, underlining to mark "punching up" (Jefferson, 2002:

[11] See Selting et al., (2009) for a description of GAT: a notation system which can be used to capture many of the same details as Jefferson's (on which GAT based), and others.

[12] The comments which follow are not intended to call into question the level of detail to which conversation analysts (or other analysts of talk-in-interaction) are able to attend. Nor are they meant to undermine the considerable heuristic value of transcriptions made using the Jefferson system of notation. The act of making and inspecting detailed transcription surely increases the chance of spotting interactionally relevant details: Jefferson (2004) gives important discussion of details with interactional relevance captured in her system and which would be (in some cases indeed were) overlooked in other, more simple transcriptions.

1379) and colons to mark prolongation of sound. Certain aspects of pronunciation can be indicated by modifications from the usual spelling of lexical items. The transcriptions are fairly straightforward to read, especially if the reader is fluent in the language being transcribed. Finally, the transcriptions are by and large suited to their purpose: they capture sufficient detail to facilitate discussion (either orally or in print) at a level suitable for many analysts in a wide range of circumstances.

As with any notation system, Jefferson's is not without its limitations. Here I discuss four issues related to the use of this system to address issues of phonetic features in interaction.

First, *do, or should, transcriptions aim to capture all phonetic features of potential interactional relevance?* Transcriptions employing the Jefferson's system of notation usually seem to be prepared in line with the mandate "put down in the transcription what you hear". It is beyond question that following this mandate has led to new findings concerning the organisation of interaction. Following this mandate also usefully places the emphasis on auditory analysis, and guards against an analyst only making reference to acoustic measures which may not be auditorily available. However, claiming to follow this mandate does raise the issue of whether everything that is auditorily available — and therefore, presumably, potentially interactionally relevant — is being captured. That does not seem to be the case. For instance, in a pair of companion articles John Kelly and John Local show that Jefferson's transcriptions do not capture all of the cases they identify of articulatory assimilation (the production of a sound at once place in the mouth in anticipation of a following sound), one interactive function of which is the local projection of more talk (Local & Kelly, 1986; Kelly & Local, 1989b). They also show that glottal closures made at the end of conjunctions and held through a silence until the start of a next word are not captured systematically (and despite the presence of a symbol for such a 'cut-off' in Jefferson's notation).

The interactional relevance of such a held closure is to hold the turn. Local & Kelly observe that where a co-participant starts up just after a 'holding silence', it is the current speaker — not the next speaker — who holds the turn: a pattern Jefferson did not identify in the paper to which Local and Kelly's work is tied

(Jefferson 1983). The patterns Local and Kelly identify arose not from 'machine readings', but from auditory analysis of the type described above.

To return to a phenomenon discussed above: one of the features of “and-uh(m)” and “but-uh(m)” which marks out that what follows will be a resumption of prior talk is its production with broadly level pitch, roughly an octave above the speaker’s baseline pitch (Local, 2004). There is no readily available means for representing the placement of talk within a speaker’s pitch range in Jefferson’s system, despite this having been shown to have interactional relevance.

Couper-Kuhlen (1996) demonstrates the interactional relevance of the relative placement of talk in a speaker’s pitch range, and in relation to the talk of a co-participant. Examining sequences from calls to a radio quiz show in which callers make an attempt to solve a riddle set by the presenter, Couper-Kuhlen shows that matching the pitch characteristics of a caller's guess in absolute frequency (i.e. without regard to its relative placement in the presenter’s range: *mimicry*) has one sort of interactional function and set of consequences (they preface rejections of the caller's guess); the presenter matching the relative placement of talk in the caller's range (*quotation*) has another, including the treatment by callers as a request for confirmation, which never occurs after mimicry.[13] Returning briefly to Fragment 3, the phonetic features that mark out a portion of JN’s talk as a parenthetical cannot be readily captured in Jefferson’s system, but clearly those sorts of features are relevant to the structuring and interpretation of the talk.

[13] Interactional phenomena, e.g. quoting and mimicry (Couper-Kuhlen 1996), back-connecting "and-uh(m)" (Local 2004) and collaborative completions (Local, 2005), have been shown to be systematic in terms of their pitch characteristics, and these characteristics have been expressed relative to speaker's baseline values. This does not necessarily mean that these, or other phenomena, are not systematic with regard to some other aspect of a speaker's range (perhaps most obviously mean, median or topline values). In addition to standard practice in the field and previous research, there are at least two reasons to consider the bottom of a speaker's pitch range as analytically and interactionally important. First, more speech is produced towards the bottom of a speaker's range than the top. In a sample of pitch range estimations for 20 adult native speakers of British and American English (10 male, 10 female), median F_0 was always lower than mean F_0, suggesting a negative skewing of the distribution of frequency values. Second, median, mean and topline values are more variable than baseline values (range for baseline values in the same sample = 10.7 ST, median = 20.3 ST, mean =19.6 ST, topline = 22.7 ST). Both of these observations suggest that a speaker's baseline is likely to be more readily available and less variable across speakers than those other points. It is therefore conceivably of greater interactional use. That said, listeners must be able to use other sorts of information to place an utterance in a speaker's range: Honorof and Whalen (2005) show that listeners can identify with a fair degree of accuracy where an utterance occurs in a speaker's range, even when an unfamiliar voice is heard out of context.

Second, *why is pitch not tracked throughout the course of an utterance?* Pitch is a distributed feature rather than a punctual one, and pitch characteristics are associated with any speech produced with vibration of the vocal folds. Transcriptions of pitch using Jefferson's notation system would seem to only be made at selected points, namely (i) at the ends of utterances or before other kinds of hiatus, and (ii) when there is markedly high or low pitch.

Third, *how is the use of upper case to be understood, given that there are at least three features which upper case would seem to indicate?* These features are (i) starts of turns at talk, (ii) proper nouns and first person singular pronouns, and (iii) "especially loud sounds relative to the surrounding talk" (Jefferson, 2002: 1380). Usages (i) and (ii) can be seen across the excerpts from Jefferson's transcriptions in this chapter (i.e. Fragments 1, 2, 4 and 5); usage (iii) can be seen in line 1 of Fragment 2. Is the initial "Y" of "YOU" (line 1) in upper case because it has characteristics of especial loudness relative to the surrounding talk or because it is the start of a turn-constructional unit (cf. the capital at the start of "Did" later in the same line)?

Fourth, *what is the status of the modifications to the orthography?* There are instances where orthography is modified, presumably to represent particular aspects of sound production where it is felt that they deviate sufficiently from the pronunciations which the standard orthography might be taken to represent. In Fragment 5, "Canterbury" (lines 1 and 3) is presumably changed from its usual spelling to capture particular pronunciational features, and to differentiate the productions from each other.

(5) Holt.SO88.II.1.3, 2:02

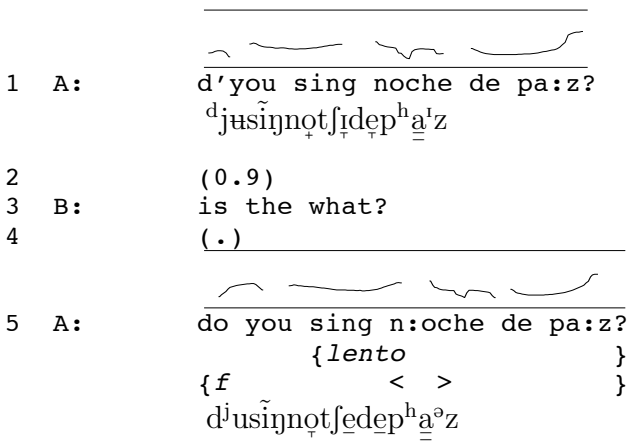
```
1  Leslie:  That's ↑near Cant'rbury isn't[i:t.
2  Hal:                                  [That's right
3           six miles f'm [Cant'[bury.
4  Leslie:                [.hhh [Ye:s. An' ↑in the↑ church you'll
5           see a lot'v my: father's family's na:me.
```

In other cases where there is an even more gross mismatch between regular spelling

conventions and the pronunciation the bulk of the original spelling is retained: "right" in line 2 of Fragment 5 is a case in point.

One way of balancing readability and detail in transcriptions has involved making a separation between the sorts of working records produced in the course of an analysis and the presentation transcriptions used in representing phonic data in print (Ball & Local, 1996). A sort of adorned orthographic method has been used in transcriptions of the latter type (see e.g. papers in Couper-Kuhlen & Ford, 2004; Couper-Kuhlen & Selting, 1996). This consists of a relatively plain orthographic transcription, capturing lexical items, some (limited) aspects of pronunciation and the sequential organisation of the talk. Analytically relevant phonetic details can then be provided on other lines. These have included e.g. voice quality, tempo, loudness (using e.g. conventions derived from music), pitch (including, perhaps, continuous representations of pitch features, or tone-labels derived from the ToBI notation set out by Beckman & Elam, 1997; Ogden, Hakulinen & Tainio, 2004 incorporate tone-labels into their orthographic transcriptions), precise details of pronunciation using the symbol stock of the International Phonetic Alphabet (IPA) and extensions (International Phonetic Association, 1999), and any other features of direct relevance to the analysis. An example of this mode of presentation is shown in Fragment 6.

(6) adapted from Curl (2005: 17)



It is worth bearing in mind that any transcription is necessarily selective: indeed, Jefferson (2004: 15) acknowledges the selective nature of her

transcriptions.[14] For example, using Jefferson's system there is no way of capturing certain features of pronunciation e.g. precise place of articulation or proximity of the articulators, consonant resonance, precise vowel qualities etc. For all of its obvious advantages, transcriptions using Jefferson's notation system will only be able to capture certain details of the original sound; on the other hand, IPA notation provides no means for capturing details of sequential organisation which Jefferson's system captures so well.

5 Future directions

This chapter has discussed some of the ways in which the phonetic design of talk figures in the management of talk-in-interaction. The tasks described here are not the only ways in which the phonetic design of talk figures in the management of talk-in-interaction. Empirical work has shown that there are other areas not discussed here where the phonetic design of talk has been shown to make a contribution e.g. the marking of stance and affect (Couper-Kuhlen, 2004a; Local & Walker, 2008), as contributing to the particular action being performed in a turn (Couper-Kuhlen, 2001; Ogden, 2006; Ogden, Hakulinen, & Tainio, 2004; Selting, 1996), and displaying an orientation to whether or not talk is fitted to the preceding talk or disjunct from it (Curl, 2005).

There are many possible future directions for work on the phonetics of talk-in-interaction. Most obviously, attempts should be made to refine our understandings of how phonetic detail functions in the management of those tasks outlined here. There are also more specialised research contexts where a combination of sequential and technical phonetic analysis could be more widely applied: see work on the talk of young children (e.g. Corrin, Tarplee, & Wells, 2001; Wells & Corrin, 2004). Clinical and other practical applications also remain under-explored (but see Auer & Ronfeldt,¨ 2004, Local & Wootton, 1995, Wells & Local, 1993). Work by Goodwin and colleagues (Goodwin, 1995; Goodwin, Goodwin, & Olsher, 2002) on interactions involving a man suffering post-stroke aphasia emphasises how a limited set of lexical

[14] A possible exception is the waveform. If a transcription is the representation of sounds on paper, then waveforms could in principle be used for this purpose. They provide a complete record of the original sound, but are clearly impractical for the sorts of tasks which we want a transcription to perform.

items can be used with different phonetic designs in the co-construction of meaning. Goodwin's work also draws attention to perhaps the most significant gap in our understanding: how phonetic resources mesh with the visual.

Since the early 1980s there has been a steady increase in the number of analysts working on phonetics and prosody in conversation, and consequently there has been a steady growth in the amount of published empirical research. There are a number of challenges which will need to be met as the field diversifies and attracts new practitioners. First, as technological developments make increasingly large databases of audio-visual material available it will be important not to lose sight of one of the core characteristics of CA: that any analysis arises out of, and accounts for, the details of single episodes of interaction. Second, relatively inexpensive but powerful computers and software make acoustic analysis readily accessible. One challenge will be to ensure that such analysis is informed by a working knowledge of (at least basic) speech production and perception. Acoustic analyses made possible by Praat and other software packages should not be seen as a substitute for careful auditory parametric analysis, but rather as its computer-based companion. Finally, it is becoming increasingly clear that there is a need to look at the speech signal in its entirety, encompassing and attending equally to features of frequency, intensity, duration and articulatory and phonatory quality as the materials to hand require.

References

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). Language in time: the rhythm and tempo of spoken interaction. Oxford: Oxford University Press.

Auer, P., & Rönfeldt, B. (2004). Prolixity as adaptation: prosody and turn-taking in German conversation with a fluent aphasic. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 171–200). Amsterdam: John Benjamins.

Ball, M. J. & Local, J. (1996). Current developments in transcription. In M. J. Ball, & M. Duckworth (Eds.), Advances in Clinical Phonetics, (pp. 51–89). Amsterdam: John

Benjamins.

Beckman, M., & Elam, G. A. (1997). Guidelines for ToBI labelling. 3rd edition. Ohio: Ohio State University. Retrieved from http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/

Bolden, G. (2009). Implementing delayed actions. In J. Sidnell (Ed.), Conversation analysis: Comparative perspectives (pp. 326–353). Cambridge: Cambridge University Press.

Corrin, J., Tarplee, C., & Wells, B. (2001). Interactional linguistics and language development: a conversation analytic perspective on emergent syntax. In M. Selting, & E. Couper-Kuhlen (Eds.), Studies in interactional linguistics (pp. 199–225). Amsterdam: John Benjamins.

Couper-Kuhlen, E. (1996). The prosody of repetition: on quoting and mimicry. In E. Couper-Kuhlen, & M. Selting (Eds.), Prosody in conversation: interactional studies (pp. 366–405). Cambridge: Cambridge University Press.

Couper-Kuhlen, E. (2001). Interactional prosody: high onsets in reason-for-the-call turns. Language in Society, 30, 29–53.

Couper-Kuhlen, E. (2004a). Analyzing language in interaction: the practice of never mind. English Language and Linguistics, 8 (2), 207–237.

Couper-Kuhlen, E. (2004b). Prosody and sequence organization in English conversation. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 335–376). Amsterdam: John Benjamins.

Couper-Kuhlen, E., & Ford, C. E. (Eds.). (2004). Sound patterns in interaction. Amsterdam: John Benjamins.

Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). Prosody in conversation: interactional studies. Cambridge: Cambridge University Press.

Curl, T. S. (2005). Practices in other-initiated repair resolution: the phonetic differentiation of ‘repetitions’. Discourse Processes, 35 (1), 1–43.

Drew, P., & Holt, E. (1998). Figures of speech: figurative expressions and the management of topic transition in conversation. Language in Society, 27, 495–522.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology, 23 (2), 283–292.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), Interaction and grammar (pp. 134–184). Cambridge: Cambridge University Press.

Fox, B. A. (2001). An exploration of prosody and turn projection in English conversation. In M. Selting, & E. Couper-Kuhlen (Eds.), Studies in interactional linguistics (pp. 287–315). Amsterdam: John Benjamins.

Goldberg, J. A. (1978). Amplitude shift – a mechanism for the affiliation of utterances in conversational interaction. In J. Schenkein (Ed.), Studies in the organization of conversational interaction (pp. 199–218). New York: Academic Press.

Goldberg, J. A. (2004). The amplitude shift mechanism in conversational closing sequences. In G. H. Lerner (Ed.), Conversation analysis: studies from the first generation (pp. 257–297). Amsterdam: John Benjamins.

Goodwin, C. (1995). Co-constructing meaning in conversations with an aphasic man. Research on Language and Social Interaction, 28 (3), 233–260.

Goodwin, C., Goodwin, M. H., & Olsher, D. (2002). Producing sense with nonsense syllables: turn and sequence in the conversations of a man with severe aphasia. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), The language of turn and sequence (pp. 56–80). New York: Oxford University Press.

Harrington, J. (2010). Acoustic phonetics. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), The handbook of phonetic sciences (2nd, pp. 81–129). Oxford: Wiley-Blackwell.

Heritage, J., & Sorjonen, M. (1994). Constituting and maintaining activities across sequences: And-prefacing as a feature of question design. Language in Society, 23 (1), 1–29.

Honorof, D. N., & Whalen, D.H. (2005). Perception of pitch location within a speaker's F0 range. Journal of the Acoustical Society of America, 117 (4), 2193–2200.

International Phonetic Association. (1999). Handbook of the International Phonetic Association. Cambridge: Cambridge University Press.

Jasperson, R. (2002). Some linguistic aspects of closure cut-off. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), The language of turn and sequence (pp. 257–286). New York: Oxford University Press.

Jefferson, G. (1981). "Caveat speaker": A preliminary exploration of shift-implication recipiency in the articulation of topic. Final report to the (British) Social Science Research Council. Report available at http://www.liso.ucsb.edu/Jefferson/ .

Jefferson, G. (1983). Two explorations of the organization of overlapping talk in conversation, 1: Notes on some orderliness of overlap onset. Tilburg Papers in Language and Literature 28.

Jefferson, G. (2002). Is "No" an acknowledgement token? Comparing American and British uses of (+)/(-) tokens. Journal of Pragmatics, 34, 1345–1383.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), Conversation analysis: studies from the first generation (pp. 13–31). Amsterdam: John Benjamins.

Kelly, J., & Local, J. (1989a). Doing phonology. Manchester: Manchester University Press.

Kelly, J., & Local, J. (1989b). On the use of general phonetic techniques in handling conversational material. In D. Roger, & P. Bull (Eds.), Conversation: An interdisciplinary perspective (pp. 197–212). Clevedon: Multilingual Matters.

Ladefoged, P. (2003). Phonetic data analysis: an introduction to phonetic fieldwork and instrumental techniques. Oxford: Blackwell.

Laver, J. (1994). Principles of phonetics. Cambridge: Cambridge University Press.

Local, J. (1992). Continuing and restarting. In P. Auer, & A. di Luzio (Eds.), The contextualization of language (pp. 273–296). Amsterdam: John Benjamins.

Local, J. (2004). Getting back to prior talk: and-uh(m) as a back-connecting device. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 377–400). Amsterdam: John Benjamins.

Local, J. (2005). On the interactional and phonetic design of collaborative completions. In W. Hardcastle, & J. Mackenzie Beck (Eds.), A figure of speech: A festschrift for John Laver (pp. 263–282). New Jersey: Lawrence Erlbaum.

Local, J., & Kelly, J. (1986). Projection and ‘silences’: notes on phonetic and conversational structure. Human Studies, 9, 185–204.

Local, J., Kelly, J., & Wells, B. (1986). Towards a phonology of conversation: turn-taking in Tyneside English. Journal of Linguistics, 22, 411–437.

Local, J., & Walker, G. (2004). Abrupt-joins as a resource for the production of multi-unit, multi-action turns. Journal of Pragmatics, 36 (8), 1375–1403.

Local, J., & Walker, G. (2005). Methodological imperatives for investigating the

phonetic organisation and phonological structures of spontaneous speech. Phonetica, 62 (2-4), 120–130.

Local, J., & Walker, G. (2008). Stance and affect in conversation: on the interplay of sequential and phonetic resources. Text & Talk, 28 (6), 723–747.

Local, J., & Walker, G. (2010). Speaking in time and sequence: phonetics and the management of talk-in-interaction. Poster presented at the Colloquium of the British Association of Academic Phoneticians. London. Poster available at http://gareth-walker.staff.shef.ac.uk .

Local, J., Wells, B., & Sebba, M. (1985). Phonology for conversation: phonetic aspects of turn-delimitation in London Jamaican. Journal of Pragmatics, 9 (2-3), 309–330.

Local, J., & Wootton, A. (1995). Interactional and phonetic aspects of immediate echolalia in autism: a case study. Clinical Linguistics and Phonetics, 9, 155–194.

Mazeland, H. (2007). Parenthetical sequences. Journal of Pragmatics, 39 (10), 1816–1869.

Nolan, F. (2006). Intonation. In B. Arts, & A. McMahon (Eds.), Handbook of English linguistics (pp. 433–457). Oxford: Blackwell.

Obeng, S. G. (1991). Pitch, loudness and turn regulation in Akan conversation. York Papers in Linguistics, 15, 221–235.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. Journal of the International Phonetic Association, 31 (1), 139–152.

Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 29–62). Amsterdam: John Benjamins.

Ogden, R. (2006). Phonetics and social action in agreements and disagreements. Journal of Pragmatics, 38 (10), 1752–1775.

Ogden, R., Hakulinen, A., & Tainio, L. (2004). Indexing ‘nothing much’ with stylization in Finnish. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 299–334). Amsterdam: John Benjamins.

Peppë, S. (2009). Why is prosody in speech-language pathology so difficult? International Journal of Speech-Language Pathology, 11 (4), 258–271.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50 (4), 696–735.

Schegloff, E. A. (1979) The relevance of repair to syntax-for-conversation. In T. Givón (Ed.), Syntax and semantics, Volume 12 (pp. 261–286). New York: Academic Press.

Schegloff, E. A. (1982). Discourse as an interactional achievement: some uses of ‘uh huh’ and other things that come between sentences. In D. Tannen (Ed.), Georgetown University Round Table on Linguistics 1981. Analysing Discourse: Text and Talk (pp. 71–93). Washington: Georgetown University Press.

Schegloff, E. A. (1986). The routine as achievement. Human Studies, 9, 111–151.

Schegloff, E. A. (1987). Analyzing single episodes of interaction: an exercise in Conversation Analysis. Social Psychology Quarterly, 50 (2), 101–114.

Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), Interaction and Grammar (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for onversation. Language in Society, 29(1), 1–63.

Schegloff, E. A. (2005). On integrity in inquiry... of the investigated, not the investigator. Discourse Studies, 7 (4-5), 455–480.

Schegloff, E. A. (2009). A practice for (re-)exiting a sequence: and/but/so + uh(m) + silence. In B. Fraser and K. Turner (Eds.), Language in life, and a life in language: Jacob Mey – A festschrift (pp. 365–374). Bingley: Emerald.

Schegloff, E. A. (2010). Some other “uh(m)”s. Discourse Processes, 47 (2), 130–174.

Selting, M. (1996). Prosody as an activity-type distinctive cue in conversation: the so-called ‘astonished’ questions in repair initiation. In E. Couper-Kuhlen, & M. Selting (Eds.), Prosody in conversation: Interactional studies (pp. 231–270). Cambridge: Cambridge University Press.

Selting, M. (2000). The construction of units in conversational talk. Language in Society, 29, 477–517.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., et al. (2009). Gesprächsanalytisches transkriptionssystem 2 (GAT 2). Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion, 10, 353–402. Retrieved from http://www.gespraechsforschung-ozs.de.

Stivers, T., Enfield, N., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences, 106 (26), 10587–10592.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. Journal of Phonetics, 35, 445–472.

Walker, G. (2004). On some interactional and phonetic properties of increments to turns in talk-in-interaction. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 147–169). Amsterdam: John Benjamins.

Walker, G. (2007). On the design and use of pivots in everyday English conversation. Journal of Pragmatics, 39 (12), 2217–2243.

Walker, G. (2010). The phonetic constitution of a turn-holding practice: rush-throughs in English talk-in-interaction. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), Prosody in interaction (pp. 51–72). Amsterdam: John Benjamins.

Wells, B., & Corrin, J. (2004). Prosodic resources, turn-taking and overlap in children's talk-in-interaction. In E. Couper-Kuhlen, & C. E. Ford (Eds.), Sound patterns in interaction (pp. 119–144). Amsterdam: John Benjamins.

Wells, B., & Local, J. (1993). The sense of an ending: a case of prosodic delay. Clinical Linguistics and Phonetics, 7, 59–73.

Wells, B., & Local, J. (2009). Prosody as an interactional resource: a clinical linguistic perspective. International Journal of Speech-Language Pathology, 11 (4), 321–325.

Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. Language and Speech, 41 (3-4), 265–294.

Wells, B., & Peppë, S. (1996). Ending up in Ulster: prosody and turn-taking in English dialects. In E. Couper-Kuhlen, & M. Selting (Eds.), Prosody in conversation: interactional studies (pp. 101–130). Cambridge: Cambridge University Press.