Phonetic Perspectives on Interaction: Ways of Observing Speech

Phonetics is the systematic study of the sounds of speech; traditionally, it considers issues such as how speech can be transcribed, matters of speech production – such as the articulations made in speaking, and the acoustic patterns they give rise to – and speech perception. Phonetics is used in this chapter for the study of the audible, outward, physical manifestation of spoken languages (i.e. not of sign languages).

Speech is the audible manifestation of language, so wherever there is speech, there is phonetics. An immediate issue when we encounter data can be not the lack of material, but where to start. One of my goals in this chapter is to explain how I go about tackling data as a phonetician. As a discipline, phonetics is increasingly exploring connections between speaking and other embodied behaviours, including gesture and facial expression (e.g. Wagner et al. 2014 for an overview) – aspects of communication which phonetics traditionally does not include (Laver 1994). This chapter takes the traditional perspective, for analytic simplicity.

If you are coming to the phonetics of talk-in-interaction from a background in phonetics, then you will be shifting your perspective towards exploring what social actions are accomplished through phonetic practices. (For an overview of the phonetics of talk-in-interaction, see Local & Walker, 2005; Ogden, 2021; G. Walker 2013; see Ogden & Walker 2013 for a discussion on the kinds of social actions that have phonetic exponents.) If you are coming to the phonetics of talk-in-interaction from a CA background, then you will be encountering a technical and systematic way of making observations of speech. The pairs of turns at talk below, where two assessments (a first assessment and a second assessment) are made (Pomerantz 1984), will help to illustrate:

27.1. Ogden 2006: 1767; gw/00.washing machine.aiff

```
01  A:  it's `BETter than `tOkens thOUgh,
02  B:  `YES it `IS bEtter than 'tOkens,
```

27.2. Ogden 2006: 1766; nrb/01.reluctant lover

```
01  A:  she's `REA:lly ^nI:ce `Isn't she.
02  B:  she `IS nIce,
03      I dO fI:nd that she just sAy:s stuff
```

```
04      just for the sake of `SAYing stuff thOUgh,
```

27.3. Ogden 2012a: 206; VT19.12.03 sea kale

```
01  A:  it's de`LICious Gail;

((two lines omitted between other speakers))

02  B:  it `!IS! delicious. hehe
```

In these transcriptions, which use GAT2 (Couper-Kuhlen & Barth-Weingarten 2011)[1], accentuation is marked with capitalisation rather than underlining. The three examples both contain the verb 'is', which in English can have either a 'strong' form, 'is', or a cliticised, 'weak' form, ''s'. In Example 1, B repeats A's words 'it's BETter', but in a slightly different form, 'it IS better'. In the first turn, there is an accent on 'better', but in the second turn it is on 'is'. In linguistic theory, this is called 'deaccenting' (e.g. Wells 2006): where a content word like an adjective is repeated, it generally does not get accented in the same way. In (1), the recycled words convey B's agreement with A's first assessment. In (2), there is a partial lexical repeat: B leaves out the modifier 'really', but the accent falls on 'is' because the adjective (assessment term) 'nice' is repeated. But 'nice' is a less positive assessment than 'really nice', and B's next action is to give a reason why 'she' might not be so 'nice'. So B displays weak agreement with A. In (3) the same assessment term is used, but 'is' is produced with a very 'strong' accent, i.e. loud and with high pitch.

For readers coming from a linguistics background, these examples show that deaccenting is one of the resources implicated in turns that somehow match (or don't) a prior turn. For those coming from a background in CA, these examples illustrate that there are quite regular and predictable linguistic patterns which help to construct the social actions of agreement and disagreement in second assessments in second position. These observations from different perspectives are complementary, and in studying the phonetics of talk-in-interaction, we take advantage of this complementarity: we explore how phonetic resources are recruited in everyday interaction in quite regular, systematic ways. In this chapter I want to show how we might come to make more such discoveries.

This chapter contains first an overview of the process of observing and transcribing, including reflection on the annotation of features that extend over longer stretches of speech, and finishing with some suggestions for how to approach data in a data session from a phonetic perspective; there is discussion of the term 'prosody', showing some of its

[1] GAT stands for *Gesprächanalytisches Transkriptionssystem*, 'conversation analytic transcription system'. It was originally designed for German.

theoretic and historic underpinnings, and discussing some common assumptions about prosody; and then the bulk of the chapter is a case study going through a short sample of speech line by line to examine some of its phonetic features and how they are used as resources for interaction. Finally, we conclude with some key messages.

27.1. Observing and Transcribing

The first word of the title of this section is 'observing': this is because all transcription depends first and foremost on observation. Observation is a skill that we can learn, practise and hone. Not only that, but developing skills of observation is entirely consistent with CA's stance of 'order at all points' (Sacks 1992 (Vol. 1): 484).

Learning phonetics takes time and dedicated practice; so this chapter cannot even scratch the surface of that: Ogden (2017) provides an introduction to the phonetics of English. What I aim to do here is to convey something of a phonetician's approach to talk-in-interaction, so the reader can appreciate the disciplinary differences between what look like the same tasks, and to provide some suggestions for further reading and exploration. The stance adopted here is one that has been developed at the University of York since the 1980s (notably in Kelly & Local 1989), combining traditional British-style phonetics, with a firm grounding in ear-training, transcription and production, with the analytic tools of phonetics more widely and CA more specifically.

27.1.1 Transcription

Transcription is a practice shared by phonetics and CA. While the practices of transcribing are similar in the two disciplines, they also take distinct approaches.

Phonetics is grounded in theories of how the vocal tract is used, and the sounds that are made in the languages of the world. In the era before recordings, phoneticians developed techniques for listening to speech, reproducing it and making phonetic transcriptions, which was in effect the only way to make records of speech. Because transcriptions are used for many different purposes, phoneticians have also theorised the practice of transcription over many years (see Heselwood 2013 for an overview).

In the 1800s, when phonetics as we now know it started to be developed and institutionalised internationally, a primary concern was illiteracy (MacMahon 2013: 116). One consequence of this was that a style of transcription was developed that was suitable for writing and reading: often called 'broad' transcription, this style of transcription tries to adhere to the principle of "one sound, one symbol" while minimising the number of characters used, and avoiding diacritics (marks that modify the value of symbols) as much

as possible. While this results in an easily read transcription, these transcriptions typically lack a lot of detail.

But phonetic transcriptions can focus on details too, and it is possible to be selective about which details get most attention. In working on a language that is unfamiliar, it is usually not possible to know which details are important linguistically, so phonetic transcriptions of this kind of data tend to be very detailed. Later work might clarify which details can be simplified or reduced, e.g. because they are predictable or consistent.

Working in a time before recording devices existed also shaped the kind of material it was possible to work with. In general, traditional phonetics (indeed, traditional linguistics) emphasises careful speech produced at first in isolation, and in monologue. Broad transcriptions in particular do not usually represent individual productions of speech; they are schematic, and usually generalise across multiple productions or many speakers, and over the particular situation of production (much like you might find in a dictionary entry). This is not the case for the data we work with in CA: people are not producing careful speech so we can transcribe it, but for some other purpose, and they only say what they say once.

Phonetic theory has plenty to say about approaches to and the uses of transcription (see, e.g. Handbook of the International Phonetic Association, 1999**;** Heselwood 2013; Laver 1994; see also G. Walker 2013, 2017b for further discussion from a CA perspective); it is not my intention here to review that literature, but it is worth reflecting that transcription serves several purposes. Phonetic transcriptions reflect the purpose they will be put to; but this means there is a responsibility on transcribers to explain the conventions and to be clear about the style and purposes of a transcript.

Because transcriptions are a way to record observations, and then in many cases to provide the raw material for analysis, there are reasons to care deeply about matters such as consistency: have the same parameters been attended to in all the data? If something is not present on a transcript, is it because it doesn't occur in the recording, or is it because the transcriber missed it?

On the other hand, transcripts are always an imperfect and incomplete product. So while we might care deeply about getting the details right, it is easy to be detached from them, and remember that they are not the data, but *representations* of the data. There are many ways to represent events or patterns in data: transcripts are one, but spectrograms, waveforms, descriptive statistics or technically precise statements of recurrent patterns are

others.[2]

CA transcripts differ from phonetic transcripts in several ways. One is straightforwardly that CA transcripts are based on orthography. For a language like English, which has very complicated sound-to-letter correspondences, this can be tricky. Another difference is that CA transcripts capture many features of talk which linguists have tended ignore as unimportant, or as errors of production: silences, gaps, repairs, hesitation markers, overlapping talk, and even the attribution of talk to speakers are all things that CA handles more formally than the International Phonetic Alphabet (IPA), which has been designed to represent the wide variety of sounds of the languages of the world in a convenient form (Handbook of the International Phonetic Association, 1999: 3).

For the purposes of readability, it is usually easier to present transcripts using conventions such as those of Jefferson, or the more linguistically-informed GAT2 system (Selting et al. 2009; Couper-Kuhlen & Barth-Weingarten 2011 for English). These both use orthography as their base, and that makes it difficult to include phonetic details. Combining IPA symbols with orthographic ones is not as straightforward as it might seem, because the same symbol might have very different uses in orthography vs. in phonetics. For example, the symbol <!> is used in orthography as an exclamation mark; in the IPA, it stands for an alveolar click sound; and it is used in the ToBI intonation transcription system to mark down-stepped tones (i.e. tones that are lower than expected), and in GAT2 (Couper-Kuhlen & Barth-Weingarten 2011) to mark an extra strong accent. The solution to this is that in presenting transcripts, it is important to be clear what the transcription conventions are.

27.1.2 Transcribing Long-Domain Voice Quality

My first project involving CA and phonetics was about creaky voice in Finnish turn-taking (Ogden 2001, 2004). Creaky voice is commonly called a 'suprasegmental' or 'prosodic' feature of talk (see Section 2). In Finnish, this is regularly towards the end of a turn at talk, and once it starts, it is either sustained to the end of the turn, or followed by another voice quality, which is either whispery, or voiceless (or both, but in that order), and then sometimes by an audible exhalation.

2 For example, two important collections of papers on prosody in conversation, Couper-Kuhlen & Selting (1996) and Couper-Kuhlen & Ford (2004) contain transcripts CA transcripts augmented with IPA transcriptions and acoustic evidence. Local (2003), a relatively early paper on the phonetics of talk in interaction in a major phonetics journal, is especially rich in this kind of illustration. See G. Walker (2013, 2017a) for more detailed consideration of best practice.

27.4. Ogden 2004: 38; Voix bulgares

```
                                      {C-}{W--}
69  P:  onks    sulle   tuttu    TÄm{mö}{nen}h
        is-QCLI 2SG-ALL familiar this-kind
        have you heard of them
70  C:  ei  o,
        NEG be
        no
```

Example 4 illustrates this, using the notation I used in the published version of the work. As can be seen, this is a multidimensional transcription, with separate lines for the orthography, gloss into English, idiomatic English translation, and at the top, voice quality, with {C} for creak and {W} for whisper above the relevant part of the orthographic line. But this is not how the IPA annotates creaky voice, whispery voice or indeed exhalation.

Creaky voice in the IPA is standardly annotated with a tilde below the relevant symbol: so the transcription of the creaky part of *tämmönen* above would be [m̰ø̰]. The standard IPA has no symbol for annotating whispery voice. This is because the IPA represents contrastive sounds in the world's languages, and so far whispery voice has not been shown to contrast with some other voice quality, although of course it is a common voice quality. To find a symbol for whispery voice, we have to go to the Extensions of the IPA and a chart called VoQS: Voice Quality Symbols. The convention there is to use a dot underneath the relevant symbol. So the end of *tämmönen* could be transcribed as [m̰ø̰n̤e̤n̤], so a transcription of the whole word could be [tamm̰ø̰n̤e̤n̤]. For exhalation there is no particular transcription convention.

While the IPA transcription is 'accurate', it is also problematic. Firstly, any reader would have to recognise the diacritics used to represent voice quality, and be comfortable reading modified letters. Secondly, the transcription is totally in IPA, so the reader would also need to recognise that IPA [a] corresponds with orthographic <ä>, and IPA [ø] with orthographic <ö>. These are not assumptions one can necessarily make. Thirdly, and more important analytically, is the fact that the extent of creak and whispery voice varies, but the order is always the same, and this ordering is not so clear from this transcription. For this reason, I decided to use another convention, taken from Extensions of the IPA (ExtIPA; ICPLA 2015, Ball et al. 2018), which is to separate voice quality from the consonants and vowels. ExtIPA actually does this using curly brackets. Following the conventions set up there, one could transcribe this word as [ta {V̰ mø V̰}{V̤ nen V̤}]. While this avoids some repetition, it is unwieldy. (It is worth noting that the GAT2 conventions offer a similar way to represent long-domain features, so that a stretch of creaky voice can be represented

<<creak> like this>.)

The solution I came to (which can be seen in Example 4) was to mark voice quality separately on another line above the orthographic line; to use the ExtIPA convention of curly brackets to represent the extent of a feature; and to represent the extent with a dashed line above the relevant stretch of talk. This has the advantage of capturing phonetic detail in a way that is clear and relevant for the current analysis, while also keeping the familiarity (and so practical simplicity) of the regular orthography. A similarly tiered approach is normally taken for multimodal transcripts (Mondada 2018).

From this it can be seen that while transcription schemes (and I have mentioned four, without discussing Jefferson's) offer solutions, these solutions may be sub-optimal, and a different solution needs to be developed. If that is done – and on the whole we should avoid the proliferation of transcription systems – then it is best to adopt a solution that has some organic relationship with other extant ones; and most importantly, the conventions for interpreting the transcription must be transparent.

27.1.3 How I Approach Data in Data Sessions

One of the joys of data sessions (see Betz, this volume) is encountering the unexpected, and being exposed to unfamiliar material. In this situation, it is often useful to have a framework to work with. As a phonetician, I have a technical framework for observing, describing and analysing speech. This does not mean that I always know how to interpret or analyse what I can observe — if I did, I wouldn't be doing research. But it does provide me with some starting points.

If I get stuck, if my mind freezes, or if I feel like my imagination is lacking, then I find that a good question to ask myself is: what is the problem that the speaker has to solve now? This may be something that reflects issues of coherence between units in talk, or possible alternative hearings, or matters of turn management. For example: 'how is the talk at this gap marked as only temporarily suspended, and not complete?'; 'there's a point of syntactic and pragmatic completion here, but the speaker continues talking; how do they resolve that?'; 'if this TCU doesn't sound coherent with the speaker's prior one, it risks sounding like something new'. Asking this question is often a good way to sharpen the analytic focus and find interesting questions.

The list below contains things that I have found are often useful ways to attend to the phonetic details of data when listening in an unmotivated way (as opposed to focusing on particular details as part of a specific project). The list is organised by starting with the turn currently of interest (the focal turn), looking at its internal features; and then comparing it

with what came before and what comes after, and considering what contribution the phonetic features make to the overall action implemented through the focal turn. (See Walker 2013 for a similar approach.) Having a framework like this makes it easier to get started; and since complicated questions can most often be broken down into several much simpler ones, it is often helpful to get an understanding of data that might seem impenetrable.

A. Features within the turn. A non-exhaustive list:

1. Long domain features of the turn such as: tempo, loudness, voice quality (including articulatory setting)[3]
2. Intonational features of the turn (for overviews, see Szczepek Reed 2010; Warren & Calhoun 2021):
    - 2.1. Phrasing: what are the intonational 'chunks' of the TCU? Many sentences can in principle be phrased in several ways, and the choice of phrasing may be consequential – see e.g. Couper-Kuhlen (2021) for a discussion of 'oh' + 'okay', and Local & Walker (2012) for discussion of issues around phrasing in syntax and phonetics
    - 2.2. Accentuation: what items are prominent and carry intonation contours?
    - 2.3. Boundaries and edges: how do the chunks/phrases start and end? See Ogden (2021) for an overview.
        - For beginnings: noises made in preparation for speaking including clicks and in-breaths; initial pitch level; initial articulatory setting (e.g. glottal stop or other sound with a rapid onset vs a more gradual beginning). See, e.g. Couper-Kuhlen (2001), Dilley et al. (1996),
        - For endings: loudness, voice quality, articulatory settings; 'trail off' vs abrupt ending; pitch level. See, e.g. Local & Walker (2004, 2005, 2012), Walker (2017a)
    - 2.4. Pitch range: does it sound like the speaker's average? Higher? Lower?
    - 2.5. Pitch span: how big does the difference sound between the highest and lowest pitches?

3 'Long domain' is deliberately unspecific; it is traditionally used to refer to anything that extends over more than a vowel or a consonant. 'Articulatory setting' refers to postures sustained in the vocal tract, such as raised or lowered larynx, nasalisation, lax or tense articulations.

3. Other features of production: how are gaps and silences entered into and exited (e.g. Local & Kelly 1986; Ogden 2001)? Are there notable features in the production of consonants or vowels (e.g. Barth Weingarten 2012, Curl 2005, Local & Walker 2004, Plug 2005)?

B. How does this turn relate to the prior one? With a systematic description of the target turn, it is possible to think about its relationship to what comes immediately before, by examining the prior turn for the same features. The 'lay person' listener has a good sense of this ("these sound alike"), but the 'phonetician listener' is able to express this similarity in more technical terms.

1. Are the turn's phonetic features **disjunctive** with the prior turn?

1.1. For example, are there noticeable differences in loudness, tempo, pitch level, or pitch span, articulatory setting or other features? Typically, the sense of 'noticeable change' is a holistic one, and we need not expect every feature to be different; but it is useful to specify what the differences are.

1.2. If this turn repeats some part of the syntax or lexis of the prior turn, how are these repeated elements produced? For example, is there deaccentuation as we saw in Examples 1-3? If not, what are the interactional consequences (e.g. producing something with the features it would have in first position, despite being in second position; cf. Heritage & Raymond 2005).

2. **Do its phonetic features match those of the prior turn?** Does it repeat or match elements of what came before? e.g. similar height in the speakers' pitch range; similar pitch span; rhythm. Matching is often a good sign of an affiliative, aligning action (Couper-Kuhlen 2012a, Zellers & Ogden 2014; cf. Ogden & Walker 2013: 293).

3. **Are there features that match what we know from other studies?**

C. How does the current turn relate to a structurally relevant prior one? Sometimes the target turn comes at some remove from an earlier one to which it is closely related structurally. A fun informal experiment is to remove the intervening material and play the two turns next to each other. Sometimes this shows an uncanny precision matching between the later and earlier turns (e.g. Couper-Kuhlen 1996; Local 2003, 2004), in itself showing that such detail is relevant to participants (at least to speakers).

Other useful prompts:

- if this is a second pair part which is displaced from its first pair part by an insert sequence, how does it relate to the first pair with which it belongs?
- if it is connected to a prior turn in some other way, how does it relate to that?

D. How does the current turn relate to the next one? Sometimes turns project particular designs in a next turn:

- If there is incomplete syntax, can a case be made for an incomplete intonational phrase too?
- Are there features of this turn that the next one picks up on?
- Is the ending of this turn rhythmical? if so, does the next turn align with the rhythm of this one? i.e. is there rhythmicity across the turn space?

E. How do all these features map on to aspects of interaction such as:

- The action being implemented in the current turn
- The sequential position of the current turn
- Aspects of speaker transition

This list is not exclusive, and it is easily adjusted to include other features or orders of organisation. What is important is that it breaks down some quite complex tasks into much smaller, more tractable ones, which have specific answers expressed in a terminology that provides technically accurate descriptions.

27.2. Prosody, Syntagmatic and Paradigmatic Relations

Before we go much further, I will say a little more about the term ‘prosody’, which I have already used but not explained; in particular something about its history and multiple meanings.

One of the foundational texts of linguistics is Saussure’s Course in General Linguistics, published posthumously in 1916 (and available as e.g. Saussure 1983; see also Anderson 2021). Saussure talks about paradigmatic and syntagmatic relations, which are fundamental ways to look at language.

Let’s start with paradigmatic relations. If we take a word like ‘mat’, it is possible to replace one of the sounds and come up with a different word: pat, bat, gnat, cat, hat, sat…

and so on. This is a basic proof procedure for showing that sounds have linguistic value: by replacing one sound with another, we make a different word. The relationship between these sounds is 'paradigmatic': one sound can be swapped for another, resulting in a different linguistic structure. The same procedure can be applied to the vowel (met, mitt, mutt…) and to the final consonant (mad, map, mac, man, mass…). This is an 'either-or' relationship, as Hjelmslev later put it (Hjelmslev, 1961; see also Anderson 2021): it's a relationship where one thing can be swapped for another, and it establishes systems of contrast.

Syntagmatic relations are about how items fit together; they are 'both-and' relations, in Hjelmslev's terms. For example, if we take a phrase like 'the bright sun', the words in English have to come in this order, and not e.g. 'Sun the bright', or 'bright the sun'. Likewise, the sounds of 'mat' come in this order, and not, for example, 'mta', or 'atm', which not only don't occur, but can't occur in English.

Both Hjelmslev and J.R. Firth, the first professor of linguistics in the UK, used the term 'prosody' to cover syntagmatic aspects of speech: those features which hold stretches of speech together, or which are particular to certain structural positions (Firth, 1948; there are also many commentaries such as Anderson 2021; Ogden, 2006b, 2012a). This might well include for example features of pitch or voice quality; but for Firth they also include features that are particular to certain places in talk.

Modern understandings of 'prosody' are rather different: the term tends to mean 'non-segmental' or 'suprasegmental', i.e. anything that is neither consonant or vowel, which are the two classes of segments in the IPA. 'Prosody' is also (erroneously) used for one aspect of this, namely intonation. Like Local & Kelly (1989), I prefer Firth's use of the term, because it is more informative. For example, in Tyneside English, the plosives [p t k] are aspirated when turn-final (Local, Kelly & Wells, 1986); aspiration is, in Firth's sense, prosodic in this position: as the Firthians could have put it, aspiration in Tyneside is a 'prosody of turn-finality'.

When listening to a piece of data, paradigmatic and syntagmatic ways of thinking are both important. Approaching data syntagmatically means thinking about its local context; how it fits with what comes before and after it. A paradigmatic approach means thinking about alternatives: how else could it have been, what else could have come in the same place.

Working with a single piece of data as we often do in CA, one is more or less forced to look syntagmatically; but once we start to get an analytic handle on the data, and — crucially — we have a collection, and/or familiarity with others' findings as reported in the

literature, it becomes easier to think paradigmatically. The paradigmatic element of analysis needs us to be familiar with the findings of the literature on the phonetics of talk in interaction, and to build a kind of internal mental library of cases.

27.3. Working through some Data

The aim of this section is to make more public some of the hitherto private thinking and 'workings-out' that go on as part of the process of analysis – something like sharing my notebook; and I'll do that by going through a short piece of data line by line using some of the techniques introduced in earlier sections. Our primarily syntagmatic perspective might make it appear that many phenomena are 'one-offs', particular to *this* moment. However, much of what we will see are recurrent practices in English, and have to be, in order to be recognisable *as* practices. Where possible, references are provided to studies with comparable examples, to provide the complementary paradigmatic angle to the analysis.

The data comes from the CallHome corpus (Canavan et al., 1997), which is also available via talkbank.org. Participants were granted a free phone call lasting about half an hour in exchange for agreeing to have their conversation recorded. The calls are recorded on two channels, and approximately ten minutes of each call is already transcribed.

27.5. CallHome en_5254.290-297.wonderful baby

Data collected by Cong Tang

```
01  B:  and you had a kid too
02  B:  so I'm just saying you're a [little more
03  A:                              [exactly
04  B:  <<p>even MORE so;>
05      (.)
06  B:  <<f> ↑-hOw> is your ↓`BAb<<cr>y> h.
07.     (0.4)
08. A:  ! <<br> hE's ↓`WONderfu:[l->=
09  B:                          [your yOUng mAn <<p,all> I should
        say.>=
10  A:  =<<f>↑^YE::s;>=
11.     =hE's gonna be `thrEE in
        ↑`JANuar[y.
12  B:          [`?un:[b e ^ l I E v a b l e;     ]
13  A:                [<<p,all>can you belIEve it,]
14  A:  [<<pp, cr> I know>
15  B:  [`nO;
16  A:  it goes <<cr>so fast> it's ridi<<cr>culous>
```

We will focus our attention on lines 6-16.

27.3.1 The End and Start of a Sequence

Lines 1-4 are the end of the prior sequence. At line 6, the first pair part question 'how is your baby' initiates a new sequence.

The turn at l.4 is quiet, low in volume and low in pitch. Conversely, the turn at l.6 is loud (<f>, *forte*), and it starts 5 semitones[4] higher, which is hearable as a pitch reset. It is worth noting that these are features of the turn beginning, and it is at the turn beginning that the relation of *this* turn to the prior turn is most of an issue. Thinking back to one of my earlier questions – *what is the problem the speaker has to solve now?* – one issue is that this turn needs to sound like something 'new' and disjunctive from what came before. So the mismatch of phonetic features between lines 4 and 6 (low and quiet vs. high and loud) is one way to accomplish that. (See Couper-Kuhlen 2001 for how this applies to one kind of new sequence; and Zellers 2013 for a study combining both CA and statistical insights into the relationship between sequential structure and pitch level.)

[4] Semitones relate to our perception of pitch. In music, an octave is 12 semitones. Acoustically this represents a doubling of frequency, i.e. the difference between 100 Hz and 200 Hz is 12 semitones, as is the difference between 400 Hz and 800 Hz: we hear this as the same interval. Semitones are therefore a practical way to talk about relations between two tones. See Ogden (2017) for more illustration.

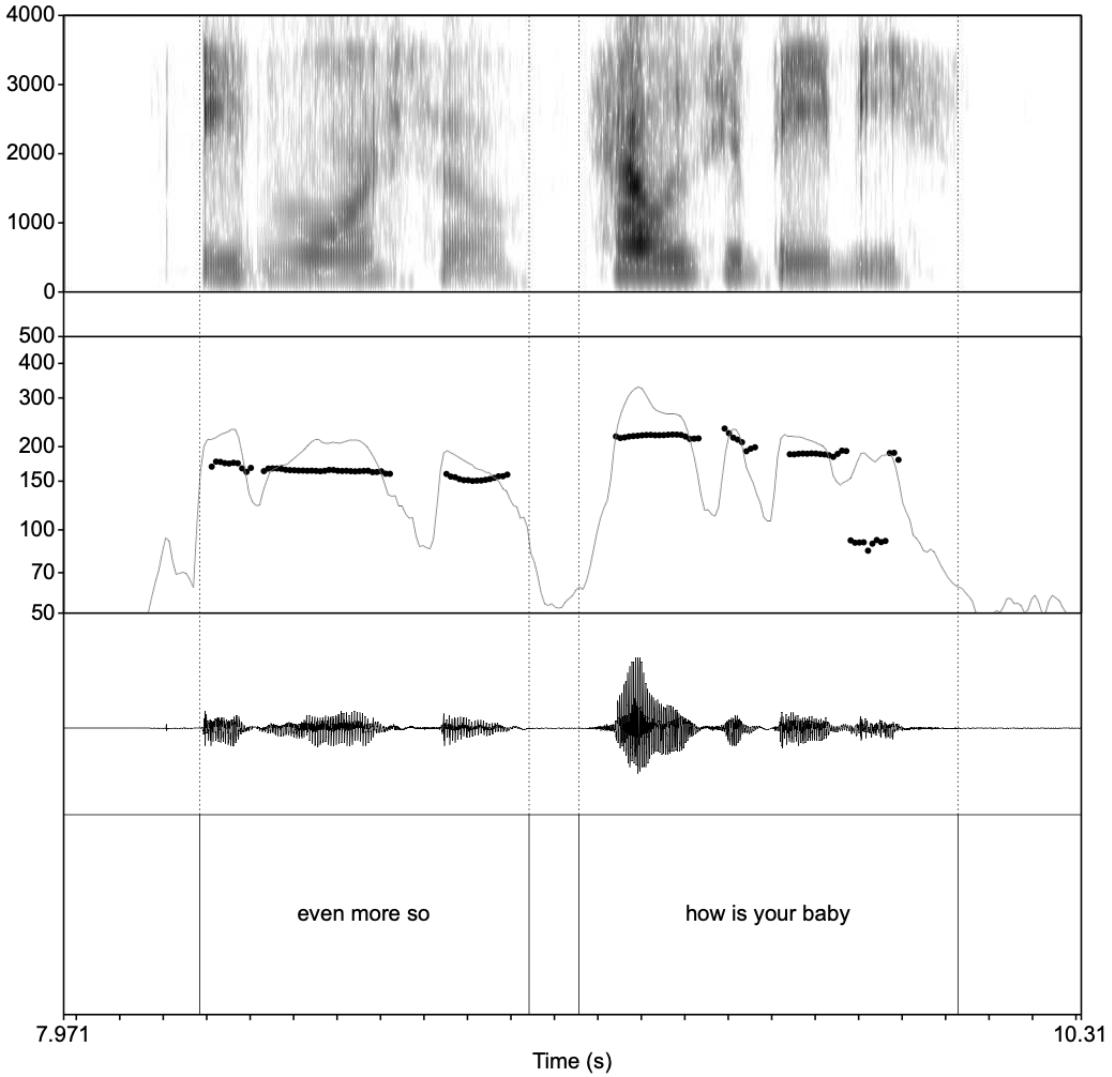

Fig. 27.1. Waveform (bottom), pitch trace (solid heavy line, middle) and amplitude (thin line, middle) and spectrogram of lines 4-6.

Fig. 27.1 shows these things. The upstep in loudness is visible in the displacement of the waveform associated with the two TCUs, and in the amplitude trace in the thin line in the middle panel. The middle panel shows F0 as thick black dots: note the upstep in F0 with 'how is...' and the creak at the end, which is visible as a low, discontinuous line at the end of the utterance.

Now notice how line 6 ends. Firstly, the focal accent in this TCU is on 'baby', which has a falling contour (ˆ in the transcript), and is produced with a downstep to a lower pitch (↓). The last syllable of 'baby' is produced with creaky voice, and the turn is followed by an audible outbreath.

I would not describe this as having 'falling pitch', which is a term that can be unhelpfully vague: it could refer to a falling contour, a phrase-final low pitch (which

usually results in a downward drift in pitch), or a downstep to a lower pitch. All three of these occur in this TCU, but they are all potentially independent. Nor would I describe the pitch here as a 'final fall', even though it falls, and it is TCU-final (see G. Walker 2017a for a deeper of turn-final intonation patterns; and Local & Walker 2012 for a discussion of other features of turn-finality). Terms like that do two things that we need to avoid: firstly, they mix form and function (i.e. there is downward movement of pitch which conveys a sense of finality), and secondly, it seems to presuppose that this is a straightforward and possibly automatic relationship between pitch movement and function (T. Walker, 2014). As a methodological starting point, it is preferable to keep the description of the phonetics separate from an account of e.g. turn regulation; this is a case that has to be made – see Benjamin & Walker (2013) for an exemplary study of how the selection of an intonation contour can be motivated. Since pitch can basically either only go up, go down or stay the same, mappings between 'ending' or 'continuation' and 'pitch movement' are overly simplistic.

In principle, there could have been an accent on 'is' instead of on 'how' and 'baby': `how IS your baby`. Producing it this way would treat talking about the baby as something that had already been ongoing, and 'how is your baby' would then be an inquiry about something already in the conversation. Produced as it is, with accents on 'how' and 'baby' is therefore another way of marking this TCU as opening a new sequence.

The creak and the exhalation are reminiscent of the Finnish data we considered in Section 1.2. There is an argument in the phonetics literature that there are regular, iconic associations between forms and functions based on physiology: in order to speak, we need to breathe out, so beginnings and endings are phonetically distinct for natural reasons. According to the Production Code (Gussenhoven 2004: 89-95) utterances begin with higher air pressure in the lungs, and therefore higher pitch and volume, whereas they end with lower air pressure, and therefore lower pitch and volume. Exhalations are a sign that there is no air in the lungs, and so they may index a temporary unavailability for talk. In this case, the exhalation comes after a very obvious first pair part, which in turn selects the other participant as next speaker.

27.3.2 Where do Turns Begin?

As Schegloff (1996) points out, and Kendrick & Torreira (2015) also demonstrate, it is not straightforward to say when a turn begins. A reasonable working understanding is that the *spoken* (or *signed;* de Vos et al., 2015) part of a turn is the 'start', since there are many possible beginnings. For spoken language, audible signs of readiness to speak, like in-

breaths, clicks, sniffs, and so on, are in pre-beginning position but do not constitute a turn by themselves (Schegloff 1996). These have been relatively understudied (but see Hoey 2014, 2020; Ogden 2013; Scobbie et al. 2011; Torreira et al. 2015), but there is one here in this extract, so we have to tackle it.

At line 7, there is a 0.4 s silence. This occurs in a place where an answer is due. We know from work in CA that a silence of this duration frequently indicates some kind of problem. Kendrick & Torreira (2015) explore ways of examining and – importantly – measuring the silence between the offset and onset of turns. They note that silences between 300 and 700 ms "may alert the recipient that the most frequent response type... is less likely" (p.287).

Alice's response in l.8 starts with a click sound ([!]) in pre-beginning position, followed by something that is recognisably an answer to Beth's question: the pronoun 'he' refers to 'your baby', and 'wonderful' targets 'how' in Beth's question.

We noted that there is a relatively long silence between the question and the answer; and now we have a turn whose first sound is not recognisable as an answer to the question. It is a click, which is one of the sounds commonly produced (perhaps incidentally) as the articulators gear up for talk.

27.3.3 Precision Timing: Rhythm across Turns

It turns out that this click is rather precisely temporally positioned, and to understand this positioning we need to discuss rhythm (see D at §1.3 above). The discussion below focuses more on the practicalities of observing and measuring rhythm in talk. For a more general overview, see Cantarutti & Szczepek Reed (2021).

Fig 2 shows the two turns, including the click, transcribed [!].

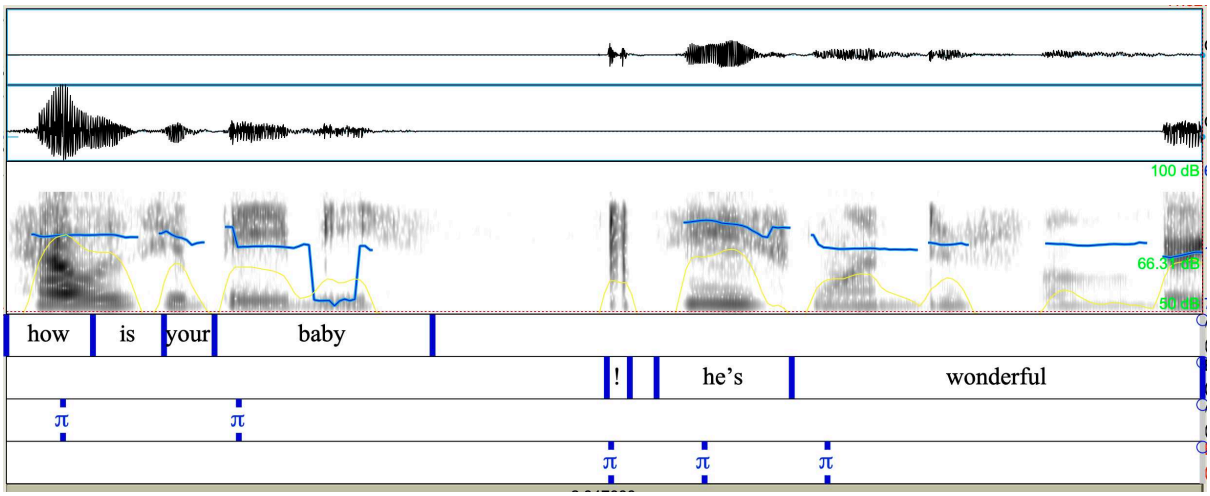

Fig. 27.2. The placement of pikes (π) – points of maximal physical activity – across two turns

The concept of 'beat' is rather common in CA, and it is used as a way to talk about rhythm and the rhythmic organisation of talk, especially as a device for projection (Auer et al. 1999; Couper-Kuhlen 1993; Schegloff 1998). The concept of beats is borrowed from music (see London 2012 for a useful overview). Beats and rhythm are percepts, and they arise from complex interactions of intensity, duration, pitch level, our knowledge of the language, and other factors too. This complexity makes it hard to mark beats in speech using acoustic criteria.

In Fig. 27.2, the label π stands for a 'pike'. This term is taken from Loehr (2007, 2012) and it refers to a **point of maximal physical activity:** events such as f0 peaks on accented syllables, eye blinks, peaks of gestures and other physically noticeable events. Pikes are a measurable way to locate 'beats' in talk (Hawkins, Cross & Ogden, 2013; Ogden & Hawkins 2015), so I will use them as a proxy for beats. The temporal interval between pikes can be measured, which allows us to compare the regularity with which pikes occur in talk; and in turn, that serves as a way to show rhythmicity more empirically.

Pikes in this discussion are annotated as follows:

- On f0 peaks (the blue/darker trace in Fig. 27.2)
- Where there is no clear f0 peak, on amplitude peaks (the yellow/lighter trace in Fig. 27.2)
- On clicks, which have no f0, but are acoustically prominent

Let's think a bit more about how these two turns in lines 6 and 8 emerge in time. Firstly, 'how' is produced with an accent, followed by another on 'baby'. There is an interval of 345 ms between these two beats. If this interval were to be repeated, then the next beat would fall approximately 345 ms later: normally ±15% deviation from the prior interval is still close enough for a next beat to be heard on beat, so anything that falls between 293 and 397 ms later is likely to sound on beat.

In the representation below, the symbol ˆ at the bottom of the transcript is used to mark the **pulse** which projects forward in time, to the temporal location of a possible next pike approximately 345 ms later.

```
              π----345----π--(~345)--
06  B:       hOw is your BAby h.
07           (0.4)
08  A:
Pulse:                              ^
```

However, Alice does not come in on this projected next beat; there is a silence, and so there is no pike. The pulse projects that the next pike after the one in this silence would occur approximately another 345 ms later, i.e. approximately 690 ms after the second pike:

```
              π----345----π--------(~690)----------π
06  B:       hOw is your BAby h.
07           (0.4)
08  A:
Pulse:                                ^           ^
```

This is not exactly what happens, but close enough: the actual interval between the pikes on 'baby' and the click is 730 ms, just 40 ms later than projected, less than 10% away from where it was projected to be, and so hearably on beat:

```
              π----345----π----------730-----------π
06  B:       hOw is your BAby h.
07           (0.4)
08  A:                                            !
Pulse:                                ^           ^
```

The adjacency pair is transcribed below, with the pikes (π) placed above the relevant part of the talk, and the intervals between them shown in ms. Alice's second pair part in l.8 has been displaced rightwards so that its temporal relation to l.6 is more obvious.

```
              π----345----π----------730-----------π
06  B:       hOw is your BAby h.
07           (0.4)
08  A:                                            ! hE's WONderfu:l
Pulse:                                ^           ^
```

Several important things can be said about this click:

- It comes in on beat. In principle it could have come in off-beat at any point from the recognisable completion of Beth's turn in 'baby'. So the mere fact of *this* temporal placement needs to be treated as having analytic significance.
- Its temporal placement displays an analysis of Beth's talk as being rhythmical enough to project a temporal slot for incoming talk: that is, the on-beat nature of the click displays a treatment of rhythm as a shared resource for handling speaker transition in time: rhythm can be demonstrated by participant orientation (Szczepek Reed 2006).

- Pre-turn clicks mark incipient speakership, and are a projection device: the click projects upcoming talk without saying anything just yet (Ogden 2013). While the click is identifiably made by Alice, it is not yet a reply to Beth's question. In other words, it displays an orientation to the relevance of a response *now*, but does not in itself constitute that response; merely a projection of a response.

Before we go further, we will just observe that although Alice's click is on beat with Beth's question, the rest of her turn is not rhythmical, and not on beat with the rhythm used across the transition space. So talk can be – in fact, normally is – rhythmical only for short stretches.

What we have seen in the transition between these turns is an example of isochrony: the intervals between beats are similar in duration. English is said to be 'stress-timed': the intervals in time between *stressed* syllables are said to be more or less isochronous. Another type of timing is syllable timed, where syllables (regardless of whether they are stressed or not) are isochronous. The literature on rhythm often classifies languages as being of one type or another. (See Arvaniti 2012; 2021; Cantarutti & Szczepek Reed 2021 for overviews.)

The view that English 'is' stress-timed is problematic, since the isochrony implied by stress-timing is not found everywhere; rather, it is found sporadically. Sometimes we notice that speech has isochronous rhythms, i.e. rhythms based on equal intervals of time between some type of unit. But often, it does not. Seen this way, rhythm is better thought of as a resource for achieving an interactional goal. In English, rhythmical incomings (which in turn are aligned with a rhythmical ending in a prior turn) seem to be used to display broadly aligning, affiliative responses to the prior turn (Couper-Kuhlen, 1993). The fact that isochronous stress-timing is sometimes present and sometimes not is what we expect if rhythmicity is a resource, rather than an omnipresent property of a language.

27.3.4 Prosodic Lamination

Having discussed the temporal placement of the onset of line 8, and the similarity of this example to others, let's now observe that the TCU in l.8 is produced with noticeably breathy voice (<br>) throughout. Breathy voice is made with incomplete vocal fold closure (Esling & Moisik 2021), and impressionistically it sounds 'softer' than modal (i.e. a speaker's regular or 'normal') voice. It is also worth observing that the final pitch is held level (the final <-> in the transcript), and is combined with a lengthening ('stretching') of 'wonderful'. Without a larger collection of similarly designed assessments, it is hard to ascribe any particular action to these phonetic features, but experimental work has shown that breathiness has positive associations (Anikin 2020).

This is a good example of the 'emotional' meanings commonly ascribed to intonation, which were discussed in Section 2.1: the voice quality intuitively seems to add something to the way in which the turn is meant to be heard, e.g. by revealing something about the speaker's stance towards what she is saying. Such cases are still poorly understood. In the phonetics literature, they are often explored with data whose ecological validity is not strong e.g. productions by actors, or decontextualised or read out materials; for example, compare Mauchand & Pell (2021) and Ogden (2010) on the phonetic design of complaints. What we can say about this turn is that its action is aligned with the action of the prior (cf. Stivers, 2008); it comes late, but its pre-beginning is on beat; and it is a positive assessment. That is not a full explanation, still less a full analysis; but it is the beginnings of one in sequential and interactional terms.

In this case, as in most others, there is no overt orientation to what 'extra' work the breathiness does, so we cannot do much more than set it aside for now and hope to build a collection of comparable cases.

There is a more general point to make here – a warning about our intuitions. One of the first ways that people interpret particularly prosodic features is along an axis of description that is broadly speaking around 'emotion'. Especially students new to CA are apt to hear something and comment that the speaker 'sounds excited', 'sounds annoyed', 'comes across as bored', and the like. This sort of gloss is normally rooted in our sense as native speakers, but as researchers using empirical methods, we need to accept it as a mere hunch, or an intuition, not an analytic response to data.

Phonetic observation leads us to well-grounded, empirical statements about the form of any piece of speech; and it treats other matters as ones for analysis. The same applies to affective displays: CA invites us to consider these as interactional, jointly achieved accomplishments (see, e.g. Couper-Kuhlen 2004, 2012a, b; Golato, 2012; Local & Walker 2008; Pillet-Shore 2012; Selting 1994 for interactional approaches to the phonetics of affective displays). Lexical choice and sequential organisation are also critical (see, e.g. Peräkylä & Sorjonen 2012). Roach et al. (1998) or Scherer (2003) provide an overview of phonetic approaches.

Acknowledging intuitions for what they are, and avoiding ascribing emotional value to the phonetic features of talk is the approach most likely to produce analytically robust insights.

27.3.5 The Phonetic Design of a TCU with Multiple Actions

The TCU in line 9 starts immediately on completion of the prior one in l.8, and it is a

self-repair on l.6, because ‘your young man’ is alternative reference for Alice’s ‘baby’. We know it is a self-repair because it is overtly marked as such with ‘I should say’.

This turn has two phonetically distinct parts. At ‘your young man’, the turn-so-far is recognisably complete, because it is a phrase that offers a replacement term for ‘your baby’, and so its status as a self-repair is clear. It is also a complete syntactic unit. The repair marker ‘I should say’, has a very different phonetic design from the first part of the turn: it is produced quiet (<p>, *piano*) and fast (<all>, *allegro*). So the turn as a whole seems to have two phonetically distinct stretches (see Kelly & Local 1989 for discussion of the term ‘stretch’).

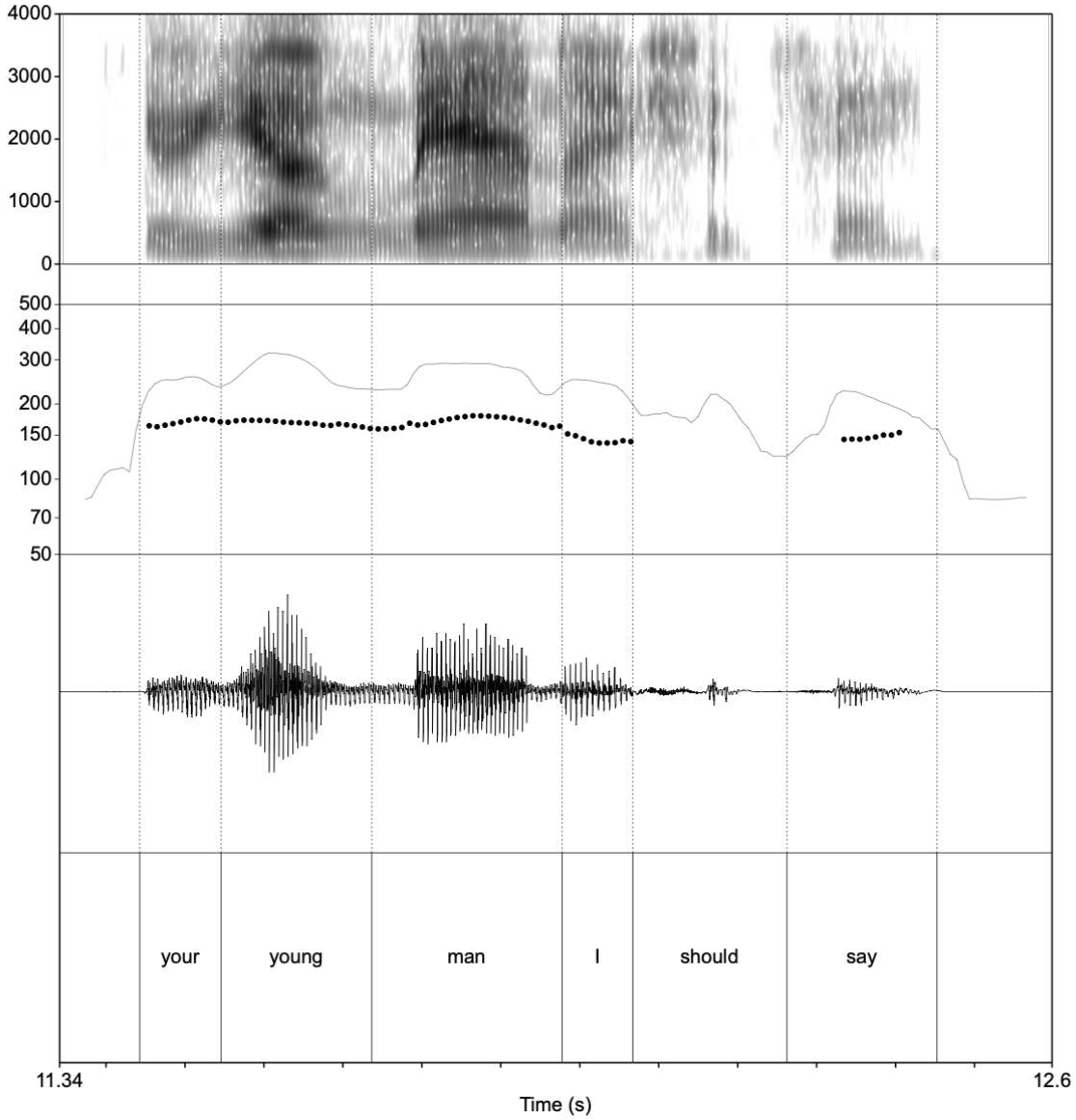

Fig. 27.3. Waveform (bottom), pitch trace (solid heavy line, middle) and amplitude (thin line, middle) and spectrogram of line 9. Note the drop in amplitude and pitch for I should say *as compared to the first half of the turn.*

This two-parted design seems to be solving the problem of marking two different actions within the turn, the first being the repaired reference, and the second being the overt repair marker, which – if it is not already clear to the co-participant – retrospectively marks the first part of the TCU, 'your young man', *as* a repair. For a similar kind of format in German, see Szczepek Reed (2015).

With this example, we see that although in principle it would be possible to produce this turn as one prosodic phrase (and as one TCU), it is phrased and produced so that there are two TCUs, each one of which marks a distinct action. We might want to make collections of examples of this kind of replacement repair (Schegloff 2013), and of this kind of overt repair marker, in order to see whether there are general practices at play here. But locally, we can see the syntagmatic relation between the two TCUs: they are related but also disjunct, both phonetically and in terms of their actions.

Lines 10-11 deal with Beth's self-repair: firstly, the 'yes' in l.10 accepts it. As with l.8, we could ask what the high pitched, loud production accomplishes; its temporal placement, which is immediately on completion of the turn in l.9, perhaps explains an intuitive sense that it is a 'wholehearted' acceptance of the repair term, where a later response might have come across as less strongly aligned. The TCU at line 11 endorses the reference to the 'baby' as 'young man', by referring to his age – evidence for the appropriateness of the new assessment, as well as why 'baby' is not the most accurate reference term.

We now have an explanation for the silence at line 7 and on-beat incoming at line 8: while Alice's response to Beth aligns with the action of Beth's enquiry 'how is your baby', there is nonetheless a problem in the formulation of the original question (i.e. 'baby' is not quite the correct reference term any longer). So while the response at line 8 is delayed, and there is a problem (which we can now see clearly), the turn in line 8 goes along with the action of the question.

27.3.6 A Prosodic Construction: Intensification

Beth's turn at line 12 consists of one word: 'unbelievable'. This in response to 'three years', which both Alice and Beth in their own ways acknowledge has being an unexpectedly long time.

This turn has features of 'intensifying emphasis' (Ogden 2012b). This is a prosodic construction (Ogden 2010) where a target word has an accent with wide pitch span, and

very elongated vowels or consonants around and just after the accent. Here, the word starts with a glottal stop (see Dilley et al. 1996 for a discussion of glottal stops in English; cf. Szczepek Reed & Persson 2016 for a discussion of glottal stops at word joins in French and German), and the nasal of the prefix 'un-' is long. The intonation contours over the rest of the word are very dynamic (though in this case with a rather narrow pitch span), with a fall over 'unbe-', then a rise-fall over '-lievable'. Examples of intensifying emphasis often come in the context of extreme case formulations (Pomerantz, 1986) or an assessment which is stronger than a prior one (Ogden 2012b). We do not exactly have that here, but we do have the age of the 'baby' given as evidence for the appropriateness of 'young man' as compared to 'baby', followed by an exchange about how quickly time passes. One of the things that intensifying emphasis does is to prosodically upgrade the meaning of what is being said: so for example, something that is 'horrible' with intensifying emphasis is more horrible than something that is 'horrible' without it; and a 'party' with intensification is a wilder, more exuberant one than something just described as a 'party' without it. Here, the details of the production of 'unbelievable' seem to orient to an understanding of time passing quickly.

We will say less about the final turns of this extract. Notice how alike they are in lexical choice; and how each participant presents both alignment and affiliation with the other. The turns are done low pitched, quiet, and there are considerable amounts of creaky voice through them. They have many of the "low and quiet" features we identified in the lines before line 6; and like those turns, these turns are sequence-closing turns.

27.3.7 Next Steps: Making a Collection

In this section we have looked at several phenomena in their natural environment. What we need to do next is to see what in these single cases is generalisable to more cases, and may be a social practice. So for a fuller analysis, we also need to work out what the generalisable patterns are, and what the system of contrast is, which means building up a collection of comparable pieces of data, including some contrastive cases.

For example, in 3.3, I showed that the click in line 8 of Extract 5 is rhythmically positioned. A next step would be to collect other turns that start with clicks. In this case, the focal turn is the response to a question, so a collection could be delimited to clicks at the start of the answer in question-answer pairs. Since question-answer pairs are one of the best studied adjacency pairs in CA, we have a good understanding of them. An analysis of clicks could consider details of positioning and rhythmicity alongside other features of question-answer pairs, like preference organisation and features of turn design. This would give us a better insight into how rhythm is deployed in English. To understand it even better, it

would be important to have examples of question-answer pairs where the transition from one speaker to another is not rhythmical and/or do not start with a click: linguistic systems are established through paradigmatic, either-or relations, so examples of other possible patterns in the same kind of position is important.

In 3.4 we noted that the turn in line 8 is produced with breathy voice. I do not have a collection of assessments made with breathy voice, but one could be made – or more usefully, a collection of assessments sorted according to phonatory setting, of which breathy would be one, and creak, falsetto or whisper could be others. Such a collection – again, combining sequential and phonetic features – could shed light on how phonatory setting contributes to the action of assessing.

In 3.5, there was a turn with two distinct actions, and two distinct phonetic designs. There are many such turns in conversation, so there are in principle many collections that could be made, depending on what the research question is. If all we want to know is whether two different actions are produced in phonetically distinct ways, then we can be eclectic in building the collection. If we want to know something more specific, like whether this design is typical of certain kinds of repairs, then the collection needs more careful circumscription.

In any case, it is important to build a collection to through both what the *forms* of the items in the collection are, their *sequential environment* and the *action* that is promoted through the relevant turn. When it comes to the forms, it is best to include contrasting items as well: for example, to understand the contribution of breathy voice to an assessment, we need to understand how assessments made with some other phonation type, or without breathy voice, work.

Work on the phonetics of talk-in-interaction has shown that phonetic design depends on sequential position and action. This is true for a speaker producing talk, and a co-participant who has to make sense of another's talk, and recognise the actions that talk promotes. For an analyst, collecting all examples of a phonetic form without paying attention to sequential position and action is likely to lead to a mass of highly intractable data which appears random, or just to show some vague statistical trends. The analyst can also look for evidence of participant orientation: in phonetic data (as in other kinds of data), this is seen through recycling, repair, or other orientation to the practice through their subsequent talk. This is how we can produce analyses that are grounded in data.

27.4. Key Messages

CA provides us with a rigorous method for making arguments about the functions of

phonetic details, not least by closely examining other orders of detail such as sequential organisation, lexical choice, and features of participant orientation; the 'linguist's theory of meaning' (Ladd, 1996) – which is little more than intuitions – can be set aside in favour of a more empirically grounded approach. This approach contrasts a traditional preoccupation with lexical contrast (Local 2003). An interactional perspective on well-known phenomena can be quite a different one from a conventional linguistic one, because in dealing with interaction, we are necessarily dealing with moments which for participants have social import (see chapter on grammar by Couper-Kuhlen et al., this volume).

Those coming to phonetics from CA face the challenge of learning how to make technical descriptions of the features of talk. Phonetic observations can be captured in technical ways, whether through theoretical concepts or measurement. Phonetic observations are an important part of the overall toolkit available to analysts. Walker (2017b) provides useful guidance on the presentation of data. Phonetic observation should be systematic; and it is bound to be partial. Measurements or labels require a systematic approach. It is the job of the analyst to explain the findings.

In working through data, it is neither necessary nor possible to account for all details at once. A good starting point is to break complex questions into simpler ones, and to focus on a limited number of things at a time. The data will still be there for analysis later.

Linguistic phenomena are often treated by linguists as objects, or the observable manifestations of an underlying linguistic order. Looking at talk-in-interaction forces us to see talk as unfolding, emergent, and evanescent; and our focus is perhaps less on 'language as object', and more on 'language as process and practice'. Interactionally oriented accounts do not necessarily negate linguistic accounts, but they often highlight the complexity, subtlety and richness of linguistic systems as vehicles for social action, and lead us to question whether linguistic phenomena are the monolithic systems they may appear to be.

References

Anderson, Stephen R. (2021). Phonology in the Twentieth Century. Second edition, revised and expanded (History and Philosophy of the Language Sciences 5). Berlin: Language Science Press. http://langsci-press.org/catalog/book/327

Anikin, A. (2020). A Moan of Pleasure Should Be Breathy: The Effect of Voice Quality on the Meaning of Human Nonverbal Vocalizations. Phonetica, 327–349. https://doi.org/10.1159/000504855

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. Journal of

Phonetics, 40(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Arvaniti, A. (2021). Measuring Speech Rhythm. In R. Knight & J. Setter (Eds.), The Cambridge Handbook of Phonetics (Cambridge Handbooks in Language and Linguistics, pp. 312-335). Cambridge: Cambridge University Press. doi:10.1017/9781108644198.013

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

Ball, M., Howard, S., & Miller, K. (2018). Revisions to the extIPA chart. Journal of the International Phonetic Association, 48(2), 155-164. doi:10.1017/S0025100317000147

Barth-Weingarten, D. (2012). Of Ens 'n' Ands: Observations on the Phonetic Make-up of a Coordinator and its Uses in Talk-in-Interaction. Language and Speech, 55(1), 35–56. https://doi.org/10.1177/0023830911428868

Benjamin, T., & Walker, T. (2013). Managing Problems of Acceptability Through High Rise-Fall Repetitions. Discourse Processes, 50(2), 107–138. https://doi.org/10.1080/0163853X.2012.739143

Canavan, Alexandra, David Graff, and George Zipperlen (1997). CALLHOME American English Speech LDC97S42. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC97S42

Cantarutti, M., & Reed, B. (2021). Stress and Rhythm. In R. Knight & J. Setter (Eds.), The Cambridge Handbook of Phonetics (Cambridge Handbooks in Language and Linguistics, pp. 159-184). Cambridge: Cambridge University Press. doi:10.1017/9781108644198.007

Couper-Kuhlen, E. (1993). English Speech Rhythm. 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(1), 29–53. https://doi.org/10.1017/S0047404501001026

Couper-Kuhlen, E. (2004). Analyzing language in interaction: the practice of never mind. English Language and Linguistics, 8(2), 207–237. https://doi.org/10.1017/S1360674304001376

Couper-Kuhlen, E. & Ford C.E. (2004, eds). Sound Patterns in Interaction: Cross-linguistic studies from conversation. Amsterdam: Benjamins.

Couper-Kuhlen, E., & Barth-Weingarten, D. (2011). A system for transcribing talk-in-interaction: GAT 2. Translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion, 12(12), 1–51. Retrieved from www.gespraechsforschung-ozs.de

Couper-Kuhlen, E. (2012a). Exploring affiliation in the reception of conversational complaint stories. In A. Peräkylä & M.-L. Sorjonen (Eds.), Emotion in Interaction. (pp. 113–146). Oxford, New York: Oxford University Press.

Couper-Kuhlen, E. (2012b). On affectivity and preference in responses to rejection. Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies, 32(4), 453–475.

https://doi.org/10.1515/text-2012-0022
Couper-Kuhlen, E. (2021). OH + OKAY in informing sequences: On fuzzy boundaries in a particle combination. Open Linguistics, 7(1), 816–836. https://doi.org/10.1515/opli-2020-0151
Curl, T. S. (2005). Practices in other-initiated repair resolution: The phonetic differentiation of "repetitions." Discourse Processes, 39(1), 1–43. https://doi.org/10.1207/s15326950dp3901_1
de Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. Frontiers in Psychology, 6(March), 1–13. https://doi.org/10.3389/fpsyg.2015.00268
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. Journal of Phonetics, 24(4), 423–444. https://doi.org/10.1006/jpho.1996.0023
Esling, J., & Moisik, S. (2021). Voice Quality. In R. Knight & J. Setter (Eds.), The Cambridge Handbook of Phonetics (Cambridge Handbooks in Language and Linguistics, pp. 237-258). Cambridge: Cambridge University Press. doi:10.1017/9781108644198.010
Firth, J. R. (1948). Sounds and Prosodies. Transactions of the Philological Society, 127.
Golato, A. (2012). German oh : Marking an Emotional Change of State. Research on Language & Social Interaction, 45(3), 245–268. https://doi.org/10.1080/08351813.2012.699253
Gussenhoven, C. (2004). The phonology of tone and intonation. New York ; Cambridge: Cambridge University Press.
Handbook of the International Phonetic Association : a guide to the use of the International Phonetic Alphabet. (1999). Cambridge: Cambridge University Press.
Hawkins, S., Cross, I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech, (September), 1–35.
Heritage, J., & Raymond, G. (2005). The Terms of Agreement: Indexing Epistemic Authority and Subordination in Talk-in-Interaction. Social Psychology Quarterly, 68(1), 15–38.
Heselwood, B. (2013). Phonetic Transcription in Theory and Practice. Edinburgh: Edinburgh University Press.
Hjelmslev, L. (1961). Prolegomena to a theory of language / by Louis Hjelmslev ; translated by Francis J. Whitfield. ([Rev. English ed.]). Madison: University of Wisconsin Press.
Hoey, E. M. (2014). Sighing in Interaction: Somatic, Semiotic, and Social. Research on Language and Social Interaction, 47(2), 175–200. https://doi.org/10.1080/08351813.2014.900229
Hoey, E. M. (2020). Waiting to Inhale: On Sniffing in Conversation. Research on Language and Social Interaction, 53(1), 118–139. https://doi.org/10.1080/08351813.2020.1712962
ICPLA International Clinical Phonetics and Linguistics Association (2015). VOQS (Voice Quality Symbols) and Extensions to the IPA. https://www.icpla.info/VOQSchart_2015v4.pdf
Kelly, J., & Local, J. (1989). Doing phonology. Manchester: Manchester University Press.
Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference : A Quantitative Study. Discourse Processes, 52, 255–289. https://doi.org/10.1080/0163853X.2014.955997
Ladd, D. R. (1996). Intonational Phonology. Cambridge: Cambridge University Press.

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

Local, J. K., Kelly, J., & Wells, W. H. G. (1986). Towards a phonology of conversation : turn-taking in Tyneside English. Journal of Linguistics, 22(2), 411–437.

Local, J., & Kelly, J. (1986). Projection and "silences": Notes on phonetic and conversational structure. Human Studies, 9(2–3), 185–204. https://doi.org/10.1007/BF00148126

Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. Journal of Phonetics, 31(3–4), 321–339. https://doi.org/10.1016/S0095-4470(03)00045-7

Local, J. K. (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: Cross-linguistic studies from conversation (pp. 377–400). Amsterdam: John Benjamins.

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. https://doi.org/10.1016/j.pragma.2004.04.006

Local, J., & Walker, G. (2005). Methodological imperatives for investigating the phonetic organization and phonological structures of spontaneous speech. Phonetica, 62(2–4), 120–130. https://doi.org/10.1159/000090093

Local, J., & Walker, G. (2012). How phonetic features project more talk. Journal of the International Phonetic Association, 42(03), 255–280. https://doi.org/10.1017/S0025100312000187

Loehr, D. (2007). Aspects of rhythm in gesture and speech. Gesture, 7(2), 179–214. https://doi.org/10.1075/gest.7.2.04loe

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. Laboratory Phonology, 3, 71–89.

London, J. (2012). Three Things Linguists Need to Know About Rhythm and Time in Music. Empirical Musicology Review, 7(1–2), 5–11.

MacMahon, M. K. C. (2013). Orthography and the early history of phonetics, (July 2022), 1–18. https://doi.org/10.1093/oxfordhb/9780199585847.013.0006

Mauchand, M., & Pell, M. D. (2021). Emotivity in the Voice: Prosodic, Lexical, and Cultural Appraisal of Complaining Speech. Frontiers in Psychology, 11(January), 1–13. https://doi.org/10.3389/fpsyg.2020.619222

Mondada, L. (2018). Multiple temporalities of language and body in interaction. Challenges for transcribing multimodality. Research in Language and Social Interaction, 51(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. Journal of the International Phonetic Association, 31(1). https://doi.org/10.1017/S0025100301001116

Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation, 29–62.

Ogden, R. (2006a). Phonetics and social action in agreements and disagreements. Journal of Pragmatics, 38(10), 1752–1775. https://doi.org/10.1016/j.pragma.2005.04.011

Ogden, R. A. (2006b). Firthian Phonology (Vol. 4, pp. 485-487 BT-Encyclopedia of Language &

Linguisti). Elsevier.
Ogden, R. (2010). Prosodic constructions in making complaints. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), Prosody in interaction (pp. 81–104). Amsterdam: John Benjamins.
Ogden, R. (2012a). Prosodies in conversation. In O. Niebuhr (Ed.), Prosodies – Context, Function and Communication (pp. 201–208). Berlin/New York: Mouton de Gruyter.
Ogden, R. (2012b). Making Sense of Outliers. Phonetica, 69(1–2), 48–67. https://doi.org/10.1159/000343197
Ogden, R. (2013). Clicks and percussives in English conversation. Journal of the International Phonetic Association, 43(03), 299–320. https://doi.org/10.1017/S0025100313000224
Ogden, R., & Walker, T. (2013). Phonetic resources in the construction of social actions. In B. Szczepek Reed & G. Raymond (Eds.), Units of Talk – Units of Action (pp. 277–312). Amsterdam/Philadelphia: Benjamins.
Ogden, R. (2017). An introduction to English phonetics. (Second Edition.). Edinburgh: Edinburgh University Press.
Ogden, R. (2021). The Phonetics of Talk in Interaction. In R. Knight & J. Setter (Eds.), The Cambridge Handbook of Phonetics (Cambridge Handbooks in Language and Linguistics, pp. 657-681). Cambridge: Cambridge University Press. doi:10.1017/9781108644198.027
Ogden, R., & Hawkins, S. (2015). Entrainment as a basis for co-ordinated actions in speech. In Proceedings of ICPhS XXVIII. Glasgow. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html
Peräkylä, A., & Sorjonen, M.-L. (Eds.). (2012). Emotion in interaction. Oxford, New York: Oxford University Press.
Pillet-Shore, D. (2012). Greeting: Displaying Stance Through Prosodic Recipient Design. Research on Language & Social Interaction, 45(4), 375–398. https://doi.org/10.1080/08351813.2012.724994
Plug, L. (2005). From words to actions: the phonetics of eigenlijk in two communicative contexts. Phonetica, 62(2–4), 131–145. https://doi.org/10.1159/000090094
Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), Structures of Social Action: Studies in Conversation Analysis (pp. 57–101). Cambridge: Cambridge University Press.
Pomerantz, A. (1986). Extreme Case formulations : A way of legitimizing claims. Human Studies, 9, 219–229.
Roach, P., Stibbard, R., Osborne, J., Arnfield, S., & Setter, J. (1998). Transcription of Prosodic and Paralinguistic Features of Emotional Speech. Journal of the International Phonetic Association, 28(1–2), 83–94. https://doi.org/10.1017/S0025100300006277
Sacks, H. (1992). Lectures on conversation, Vols. 1 & 2, edited by G. Jefferson. Oxford: Blackwell.
Saussure, F. de. (1983). Course in general linguistics; ed. by Charles Bally and Albert Sechehaye with the collaboration of Albert Riedlinger; tr. and annotated by Roy Harris. London: Duckworth.
Schegloff, E. A. (1996). Turn Organization: One Intersection of Grammar and Interaction. In E. Ochs,

E. A. Schegloff, & S. A. Thompson (Eds.), Interaction and Grammar (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (1998). Reflections on Studying Prosody in Talk-in-Interaction. Language and Speech, 41(3-4), 235–263. https://doi.org/10.1177/002383099804100402

Schegloff, E. A. (2013). Ten operations in self-initiated, same-turn repair. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), Conversational Repair and Human Understanding (pp. 41–70). Cambridge University Press.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40, 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Scobbie, J. M., Schaeffler, S., & Mennen, I. (2011). Audible aspects of speech preparation. In ICPhS XVII (pp. 1782–1785). Hong Kong.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., … Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion, 10, 353–402.

Stivers, T. (2008). Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation. Research on Language & Social Interaction, 41(1), 31–57. https://doi.org/10.1080/08351810701691123

Szczepek Reed, B. (2006). Prosodic Orientation in English Conversation. London: Palgrave Macmillan.

Szczepek Reed, B. (2010). Analysing conversation: an introduction to prosody. Basingstoke: Palgrave Macmillan.

Szczepek Reed, B. (2015). Managing the Boundary Between “Yes” and “But”: Two Ways of Disaffiliating With German ja ‘aber’ and ‘jaber’. Research on Language and Social Interaction, 48(March), 32–57. https://doi.org/10.1080/08351813.2015.993843

Szczepek Reed, B., & Persson, R. (2016). How Speakers of Different Languages Extend Their Turns: Word Linking and Glottalization in French and German. Research on Language and Social Interaction, 49(2), 128–147. https://doi.org/10.1080/08351813.2016.1164405

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. Speech Communication, 57, 209–232. https://doi.org/10.1016/j.specom.2013.09.008

Walker G (2013) Phonetics and prosody in conversation. In Sidnell J & Stivers T (Ed.), Handbook of Conversation Analysis Wiley-Blackwell

Walker, T. (2014). Form ≠ Function: The Independence of Prosody and Action. Research on Language & Social Interaction, 47(1), 1–16. https://doi.org/10.1080/08351813.2014.871792

Walker, G. (2017a). Pitch and the Projection of More Talk. Research on Language and Social Interaction, 50(2), 206–225. https://doi.org/10.1080/08351813.2017.1301310

Walker, G. (2017b). Visual Representations of Acoustic Data: A Survey and Suggestions. Research on Language and Social Interaction, 50(4), 363–387. https://doi.org/10.1080/08351813.2017.1375802

Warren, P., & Calhoun, S. (2021). Intonation. In R. Knight & J. Setter (Eds.), The Cambridge

Handbook of Phonetics (Cambridge Handbooks in Language and Linguistics, pp. 209-236). Cambridge: Cambridge University Press. doi:10.1017/9781108644198.009

Wells, J. C. (2006). English intonation. Cambridge: Cambridge University Press.

Zellers, M. (2013). Prosodic variation for topic shift and other functions in local contrasts in conversation. Phonetica, 69, 231–253. https://doi.org/10.1159/000353445pagect23

Zellers, M., & Ogden, R. (2014). Exploring Interactional Features with Prosodic Patterns. Language and Speech, 57(3). https://doi.org/10.1177/0023830913504568