
Article:
Walker, G. (2015) Phonetic Variation and Interactional Contingencies in Simultaneous Responses. Discourse Processes. ISSN 0163-853X

https://doi.org/10.1080/0163853X.2015.1056073

Phonetic variation and interactional contingencies in simultaneous responses

Gareth Walker
University of Sheffield

School of English
University of Sheffield
Jessop West
1 Upper Hanover Street
Sheffield
S3 7RA

Tel: +44 (0)114-222-0238
Fax: +44 (0)114-222-8481

Email: g.walker@sheffield.ac.uk

Running head: Simultaneous responses

Phonetic variation and interactional contingencies in simultaneous responses

Abstract

An auspicious but unexplored environment for studying phonetic variation in naturalistic interaction is where two or more participants say the same thing at the same time. Working with a core data-set built from the multimodal Augmented Multi-party Interaction (AMI) corpus. The principles of Conversation Analysis are followed to analyse the sequential organisation of the talk and to explain the phonetic variation observed. Acoustic divergence and equivalence between simultaneous responses are described. Phonetic features discussed include duration and timing, pitch, loudness and phonation type. The interactional factors which explain the acoustic divergences are established through turn-by-turn analysis and consideration of gaze direction and other visible features. It is argued that any research on phonetic variation in naturalistic talk which disregards the local organisation of interaction will always be incomplete.

Phonetic variation in simultaneous responses

1 Introduction

Accounting for the different pronunciational shapes that utterances take is a perennial challenge in speech research. Aspects of phonetic variation have been related to such factors as age, gender, social status, language background, emotional state, attitudinal stance, pathology, speaking context and so on. Different phonetic forms may be used in perceptual testing to establish one kind of relevance of phonetic variation to hearers. Another kind of relevance of phonetic variation is where it intersects with the local organisation of interaction. However, even in sociolinguistic studies of phonetic variation - the bedrock of which is spoken discourse - consideration of interactional factors is likely to be highly selective at best (Hay & Drager, 2007, p. 95). In this paper aspects of phonetic variation in spontaneous speech are related to the organisation and management of interaction. Some ways in which phonetic variation is interactionally driven are set out. Differences in the pronunciational shapes of utterances are shown to arise from participants' management of local interactional contingencies, including differences in knowledge, stance and recipiency. This has an important practical implication for studies of phonetic variation: phonetic variation in naturalistic speech must be inspected for possible connections with the local organisation of interaction. This investigation has relevance for analysts of discourse, too, especially those working in Conversation Analysis (CA) and allied areas of research (Sidnell & Stivers, 2013): a wide range of phonetic details is considered, analysed and represented which yields a richer account of pronunciational details than is typically provided by CA-based research, where the handling of phonetic details is usually more restricted and selective.

An auspicious but unexplored environment for studying phonetic variation is where two or more participants say the same thing at the same time. To date, studies of simultaneous speech have focussed on speech delivered in laboratory conditions, with subjects either reading a prepared script or shadowing what is played to them on a recording (Cummins, 2002, 2003, 2009; Cummins, Li, & Wang, 2013; Marslen-Wilson, 1985). A prototypical example of simultaneous speech in spontaneous, unscripted interaction is shown in (1). The example is taken from an unscripted interaction in which four participants are sitting around a table (three on the floor, one on an armchair) eating dinner. Except for the inclusion of screenshots, the transcription excerpt is presented unchanged from an original transcription in circulation among researchers and follows conventions widely applied in Conversation Analysis (Jefferson, 2004). The simultaneous response of interest is presented in bold text and this convention is followed throughout. Screenshots are aligned with the transcription. The symbols ∇ and Δ show where screenshots were taken relative to the transcription beneath and above the screenshot respectively. Vivian and Shane have been talking about a lunch they had together on the day of the recording, and in particular about a half lobster Shane ordered as part of a lunchtime special. Michael asks about the cost of Shane's lunch at line 1; Shane's talk at line 2 relates to talk before Michael's enquiry.

(1) Chicken dinner, 08:49
Left to right: Vivian, Shane, Nancy, Michael.

▽

```
1  Michael:    How m-How mu:c[h.
2  Shane:                   [th'body
3                   (0.5)
4  Vivian:     Si[x ni] nedy ]five.]
5  Shane:       [Six ]ninedy]five.]
```

△

At lines 4-5, Vivian and Shane respond simultaneously to Michael's enquiry in line 1. Several observations can be made about the simultaneous talk which sets it apart from some other instances of overlapping speech.

1. *The simultaneous responses are conditionally relevant responses to the same turn* (Schegloff, 1968). This is evident from the format of the responses which satisfy the constraints on relevant participation put in place by Michael's turn. There is further visible evidence that Vivian and Shane are responding to Michael's enquiry. At the end of Michael's enquiry at line 1 Shane and Vivian are looking down. By the time they they speak they have both turned to Michael. As visible evidence that she is not treating Michael's question or Shane and Vivian's responses as being addressed to her, Nancy continues to look down.

2. *The simultaneous responses begin at, for all practical purposes, the same time.* While Shane starts just after Vivian, Shane is starting up within what Jefferson (1986, pp. 164-167) proposes as a 'blind spot'. In such a 'blind spot' participants (here, Shane and Vivian) are in speakership orientation rather than monitoring the ongoing talk.

3. *The simultaneous responses are composed of the same words and syntax.* They are not concurrent and connected productions with different formats (C. Goodwin & Goodwin, 1987). It may not seem surprising Vivian and Shane produce turns matched in this way: after all, the lobster lunch presumably did cost $6.95. However, the choice of the same words and syntax is not inevitable: either speaker could have picked another format which would have satisfied the constraints placed on them by Michael's enquiry ("seven dollars", "seven bucks", "less than seven dollars/bucks", "not even seven dollars/bucks" and so on).

4. *The simultaneous responses are smooth productions and are not turn-competitive* (French & Local, 1983; Wells & Macfarlane, 1998).

5. *Either of the simultaneous responses would have been sufficient.* There is nothing in the enquiry to suggest that Vivian and Shane are both required or expected to respond. Michael's

enquiry is not designed to engender ensemble-type participation (Lerner, 1993) or any other form of talk expected to be done simultaneously rather than serially (Schegloff, 1995/2000). Both participants have the rights and knowledge to respond to Michael's question: while Nancy has been the principal storyteller Shane has contributed to the story, and the story is specifically about his lunch.

6. *The simultaneous responses each constitute a complete turn-constructional unit and a possibly complete turn.* This is as opposed to preemptive turn completions or otherwise collaboratively built turn-constructional units (TCUs) (Hayashi, 2013; Lerner, 2004, and references therein; Local, 2005; Sacks, 1992; on TCUs see Clayman, 2013).

In this paper 'simultaneous response' is used to refer specifically to responses where 1. to 6. above apply. This is the sort of simultaneous talk which, in some cultures at least, can initiate a jinx or other ritual (Roud, 2010, pp. 378-379) and the tradition of jinxing speaks to the recognisability and familiar nature of the phenomenon under examination.

Simultaneous responses afford analysts valuable opportunities to study two or more versions of the same utterance, produced in the same interactional context. Any differences between the responses can be considered against local interactional contingencies. This paper will show some of the ways in which details of simultaneous responses, and particularly differences between them, can be accounted for by a careful consideration of the local interactional environment. In doing so, this paper will expose some of the ways in which phonetic details are interactionally shaped. A key implication of this finding is that local interactional contingencies must be taken into account in dealing with phonetic variation in interactive talk. This paper has a significance for analysts of discourse beyond the findings it reports concerning simultaneous speech. To date, accounts of phonetic details in CA-based research have mostly been mostly of the 'phonetics of X' type, reporting on investigations into the phonetics of agreement and disagreement, talk- and turn-projection, overlap, repair, repetition, and so on (see Walker, 2013 for an overview). This paper does not arise from an investigation of that type into the 'phonetics of simultaneous responses' since there is no straightforward interactional warrant for such an investigation: there is no evidence that the participants are offering their talk as anything other than solo productions. This paper reports on details of talk which happens to be simultaneously, rather than serially, produced and which there is every reason to believe will figure in the management of other sorts of sequences. In other words, none of the phonetic features identified in what follows are part of the 'phonetics of simultaneous responses': rather, they reflect more general resources available for the management of talk-in-interaction.

Section 2 presents an overview of the data and methods used in the remainder of the paper. This paper presents integrated sequential and phonetic analyses of cases of simultaneous responses. Coverage will be given to responses which are acoustically divergent (section 3) relating those divergences to interactional contingencies, before dealing with a contrasting case of acoustic equivalence (section 4). Section 5 summarises the paper.

2 Data and methods

Depending on the nature of the recording it can be difficult to analyse simultaneous speech, and tasks as basic as distinguishing the voices become difficult. Generally speaking, observations have to be restricted to broad features of pitch, loudness and rate as it is often not possible to make judgments over more fine-grained features. Single-channel audio

recordings as in (1) afford the analyst no opportunity to separate out the voices and conduct the sort of technical phonetic analysis research has shown to be warranted. Single-camera video recordings (again, like (1)) can miss details captured when multiple cameras are used.

This study examines sequences from the Augmented Multi-party Interaction (AMI) corpus (http://groups.inf.ed.ac.uk/ami/corpus). The AMI corpus is described in detail in Carletta (2007). Briefly, the corpus contains recordings, transcriptions and annotations of 100 h of multi-party meetings. The recordings were made using audio tracks from multiple headset microphones and video tracks from multiple cameras. Most of the meetings involve design teams made up of four participants. Most are so-called 'scenario meetings' where fictitious teams are set the task of designing a new remote control; the remainder are 'non-scenario meetings', i.e. meetings which would have occurred anyway. According to the documentation distributed with the corpus almost half of the 187 participants for whom a native language is recorded give English as their native language; the remainder take various languages as their native.

The technical advantage of using a multi-modal, multi-channel corpus such as AMI in this context is clear: the recordings allow detailed inspection of a much wider range of phonetic and visible features than single-camera (or audio-only), single-channel recordings. While the data are naturalistic, rather than natural or naturally occurring, the instances located in the AMI corpus exhibit the same sorts of sequential and phonetic features as those in recordings such as (1). The observations set out for the simultaneous responses in (1) were used as criterial features in building the data-set for this study. There is no reason to think that the simultaneous responses collected from the AMI corpus are affected by the context in which the recordings were made in terms of their occurrence or their details.

Instances of the target phenomenon were identified through preliminary computer searching of AMI annotations using a script written by the author, followed by case-by-case verification. A computer script identified simultaneous words presented with an initial upper case character in the AMI annotations: AMI transcribers were told to use upper case characters for the start of a new sentence (Moore, Kronenthal, & Ashby, 2005), and thus, roughly, the beginning of a new TCU. To try to get at full TCUs minimal responses and connectives (e.g. "um", "uh", "and", "but", "so", "or") were only included where the item was immediately followed by a further simultaneous word. The resulting instances were then verified by listening to the extracts and comparing them against the features set out in section 1. Some latitude was permitted in the lexico-syntactic make-up of responses: for instance, in (3) one of the simultaneous responses includes a definite article whereas the other does not. Instances were excluded where necessary. Frequent reasons for exclusion were the production of matched items as a collaborative completion of an ongoing TCU rather than a new TCU, production as part of different TCUs and consecutive rather than simultaneous production.

Responses beginning within 200 ms of each other were considered simultaneous start-ups rather than the later start-up being responsive to the earlier one. (To put this duration in context, in CA transcriptions the symbolisation (0.2) indicates a silence of approximately 200 ms, and a symbolisation of (.) one of approximately 100 ms. For convenience, and since the periods of time are so small, individual measures stated in milliseconds will be rounded to the nearest 10 ms; aggregate measures, such as averages, will be stated without such rounding.) There are good grounds for considering responses beginning within 200ms of each other to be simultaneously, rather than serially, produced. If the responses began within 200 ms of each other it is assumed that the later response is starting up within the 'blind spot' proposed by

Jefferson (1986) during which the simultaneous talkers are in speakership orientation rather than monitoring the ongoing talk. The existence of such a 'blind spot' of around 200 ms is supported by experimental work: on the basis of laboratory experiments Fry (1975) showed that the mean time for subjects to react to the onset of a simple speech stimulus with a simple vocal response (thus factoring out complex processing and planning) was 213 ms. (See also Magyari, Bastiaansen, de Ruiter, & Levinson, 2014 on anticipating and reacting in conversation.) The 200 ms threshold is lower than the time it took even the best shadowers to copy ongoing talk in an experimental context (Marslen-Wilson, 1985).

Thirty-two simultaneous responses were identified. Given the complexity involved in presenting analyses of extracts in sufficient detail, a small set of these cases is presented here. Four cases are discussed in detail in this paper, reporting on detailed phonetic and sequential analysis following the approach and methodological principles set out in Local and Walker (2005). The phonetic analysis combines auditory and acoustic phonetic techniques; analysis of the sequential organisation of the talk follows the principles of CA. The cases come from three different interactions. The cases were chosen because they demonstrate especially plainly the interactional relevance of different pronunciational shapes.

3 Acoustic divergence in simultaneous responses

This section deals with cases where there are notable acoustic divergences between the simultaneous responses. It can be inferred from the transcription in (1) that in various respects Vivian's and Shane's responses are acoustically similar: both are produced with final falling pitch and both are produced with a similar overall speaking rate. There are certain differences between Vivian and Shane's responses indicated, too, most notably in Shane's response being slightly louder than Vivian's and beginning slightly later. In some cases acoustic divergences between simultaneous responses are quite small but are readily perceptible. They encompass a range of phonetic parameters. Of course, some differences in responses arise from the language background and physical characteristics of the speakers. For this reason pitch traces will be normalised to reflect the pitch ranges of the speakers, and nothing will be made of certain articulatory differences between the utterances which can be attributed to speakers' pronunciational norms. Whether or not simultaneous responses exhibit acoustic divergence, and of what kind(s), or acoustic equivalence was determined through qualitative phonetic analysis combining techniques of auditory and acoustic analysis rather than being determined by, for example, the application of computer algorithms. The descriptions of individual cases will hopefully make clear what constitutes divergence and equivalence in this context.

A case of acoustic divergence between simultaneous responses is shown in (2). The participants have been discussing movies to be screened at a movie club, with each participant proposing movies for consideration. Up to line 10 there is a move towards the closure of the meeting. (Codes at the start of the transcription excerpts identify the relevant interaction in the AMI corpus and a time in seconds for the relevant excerpt. Some portions of audible breathing, which are frequent in recordings made with headset microphones, have been omitted to increase legibility of the transcriptions. Where a frame appears around screenshots this shows that they were taken from different cameras at the same time, as in lines 11 and 12.)

(2) IB4011-2384

First screenshot, left to right: David, Ben, Adele, Chris.

```
 1  Ben:   let's cross the bridge when we g[et to i]t[hhmhh
 2  Dav:                                    [.ptehh ] [.hh okay
 3               (0.5)
 4  Dav:   .hhhhhhh
 5  Ade:   oka[y
 6  Dav:      [thank you
 7  Ben:   .hhh[hh
 8  Chr:       ['kay
 9               (0.4)
10  Ben:   thanks (.) see you A[pril nun-            twen]ty ninth
11  Dav:                       [so what's the movie again]
                                                      △
```

```
                                                      ▽
12  Ben:   Ame[rican  Beauty]
13  Ade:      [American Beau]ty=
               △
```

```
14  Chr:   =[Ameri]can Bea[uty    fi]nally
15  Dav:    [okay ]       [thank you]
16               (1.0)
17  Chr:   okay
```

Amidst the move towards the closure of the meeting David seeks confirmation of which movie has been selected ("so what's the movie again", line 11). After David's request for confirmation Ben and Adele respond with "American Beauty" (lines 12-13). These responses satisfy the criteria set out above for consideration as a simultaneous response. The response is treated by David as sufficient ("okay...thank you", line 15). The two responses start within 200 ms of each other and can therefore be considered to start simultaneously. The relative timing of the simultaneous responses in (2) is shown in Figure 1 which contains annotated waveforms of the responses.

===== FIGURE 1 ABOUT HERE

Time runs left to right on the x-axis, and the horizontal alignment of the two waveforms captures their relative timing. *S* and *E* indicate the start and end of the responses respectively; other italicised lowercase labels identify the same acoustic landmarks elsewhere in the responses. Acoustic landmarks are speech events which can be readily identified in the simultaneous responses, auditorily and visually by means of a computer display. Such

landmarks are a useful tool for describing the relative timing of the utterances, and have proved to be useful in the study of synchrony in speech (Cummins, 2002, 2006). In Figure 1, *a* marks the start of the underlined vowel in "Am<u>e</u>rican"; *b*, *c* and *d* mark the release of the plosives in "Americ<u>a</u>n <u>B</u>eaut<u>y</u>".

Adele's response in (2) begins 170 ms after Ben's: see Figure 1. The two responses thus fulfil the criterion for consideration as a simultaneous start: they are close enough for Adele's response to be considered independent of, rather than responsive to, Ben's. Nevertheless, the difference in start times represents a notable and audible acoustic divergence. The explanation for the acoustic divergence is revealed by careful consideration of the video recording of the segment, and the relationships between what can be seen (and what the participants are seeing) and the talk being produced.

At the end of the enquiry at line 11 of (2), Ben and David each have their head and gaze directed to each other. Adele is looking at David. Ben is still looking at David when he begins his response at line 12. Between the end of the enquiry at line 11 and the beginning of Ben's response there has been a significant change in Adele's head and gaze direction: she has begun to look down at the papers on the table in front of her. (Note a similar change in head and gaze direction from Chris.) These are papers on which Adele has been recording various things as the meeting has progressed, including systematically documenting each vote cast for the movies nominated to be screened. It is evident from the timing of Adele looking to the papers relative to the enquiry that these papers are relevant to the production of her next turn: the change in her gaze direction occurs just after David's enquiry and just before she responds to it. Ben is able to produce his response without looking at his notes, and can thus begin his response to David's enquiry earlier than Adele. The acoustic divergence whereby Adele starts her response after Ben can therefore be ascribed to differences in knowledge, those differences having a bearing on their capacity to respond as soon as the enquiry is issued. These differences in knowledge are only revealed by careful consideration of the video data alongside the audio.

In (3) the beginnings have a higher degree of synchrony than in (2), but there is acoustic divergence at the end of the responses: the final sound of one the responses is considerably longer than the other. The participants in (3) are members of a team which has to design a remote control for a television. Alex has been assigned the role of project manager, Bethan is the user interface designer, Carl is the industrial designer and Dale is a marketing expert. This first meeting, led by Alex, has moved into its closing phase with Alex outlining tasks for each member of the team. He is standing at a large screen showing the slide in Figure 2. Alex has set out the task for the marketing expert (the third point under 'individual actions'); in the transcribed excerpt he continues to work up the list on the slide.

===== FIGURE 2 ABOUT HERE

(3) TS3005a-1183

First screenshot: Alex (standing); clockwise from top: Bethan, Carl, Dale (all seated).

▽

1 Ale: the user .hh interface designer will (.) work out the technical functions
2 design .hhh and (.) this was the (.) interface designer or
3 Bet: mm
4 Ale: the [intera][ctio]n designer or what w[as it][aye dee]
5 Car: [mm] [no::]

▽ ▽

6 Bet: [mm] [**the in**][**terface**]
7 Dal: [**interf**][**ace: :]:**

△ △ △

8 Ale: interface designer (.) okay- first guess was right .hhhhhh uh (.) will take a
9 look at (.) the the working desig[n
10 Car: [no the indu[strial desi]gner will
11 Bet: [no the-]
12 Car: t[ake a l]ook at the working design and the:- in- uh usa[bilit]y interaction
13 Bet: [yeah] [user]
14 Ale: industrial designer
15 (.)
16 Car: [yeh]
17 Ale: [oka]y
18 Car: oka[y mmhmhhh
19 Ale: [sorry
20 Ale: mhhh hmhm
21 (0.4)
22 Car: okay [.hm hhh
23 Ale: [.hh hh .hh let's just use the acronyms

The referent of the demonstrative pronoun "this" in line 2 is the acronym "ID". It is evident from Alex's talk transcribed in lines 2-4 that he is having difficulty in providing the definition for that acronym. He first produces a candidate meaning for the acronym ("interface designer") with a trail-off "or" (Walker, 2012). In response Bethan produces "mm" with rising pitch (line 3). This is followed by Alex providing a further candidate meaning ("interaction designer") before providing an overt statement of the problem ("what was it aye dee", line 4). Bethan and Dale respond with versions of " (the) interface" at lines 6-7. At the same time Carl responds with "no" (line 5) to signal that neither of the candidate meanings put forward by Alex is correct. While Carl's right to respond to Alex's turn at line 4 might be

privileged since it is his role being discussed, Bethan and Dale's rights to respond are equal with one another. (It is not possible to tell exactly where Alex is looking though he has his head oriented towards the seated group and could conceivably have his gaze directed at any of the participants.) The solution to the repair provided by Bethan and Dale is accepted by Alex through repetition at line 8. (As might be anticipated given Carl's response at line 5 it turns out later that Alex's candidate meanings for the acronym and thus Bethan and Dale's repair solutions are incorrect and "ID" stands for "industrial designer": see lines 11 on. That Alex, Bethan and Dale have still got it wrong does not have a bearing on the deployment, design or treatment of the simultaneous response by Bethan and Dale.) Figure 3 shows the relative timing of the simultaneous responses in (3).

===== FIGURE 3 ABOUT HERE

The two responses begin within 70 ms of each other, so their beginnings have a higher degree of synchrony than those in (2). Throughout the response, up to the production of the final fricative, Dale is always ahead of Bethan with a lag between acoustic landmarks *a* to *e* (the start of the final fricative) of between 170 ms and 220 ms (mean = 192 ms). Much of this lag is due to Bethan producing an initial definite article which Dale does not produce. The duration of the final fricatives ("interface") is a key acoustic divergence between the two responses. The final fricative (from *e* to *E* in Figure 3) is much longer in Dale's response (490 ms) than in Bethan's (150 ms). Crystal and House (1988) report a mean duration of 130 ms for phrase final [s], which supports the conclusion that Dale is prolonging his final fricative rather than Bethan producing a short final fricative. Dale's final fricative can be compared with his preceding fricative in the word ("interface"). Dale's earlier fricative is much shorter, with the portion from the beginning of the fricative to the start of the following vowel measuring 120 ms. This is further evidence that Dale is producing a final fricative which is longer than might have been expected based on his talk up to that point.

It is clear that Dale's final fricative in line 7 of (3) is prolonged. What needs to be established is what, if anything, this prolongation allows him to achieve. As a preliminary to that, it is necessary to provide an informed estimate of where his talk would have ended if his final fricative had not been prolonged. Such an estimate will allow a comparison of how things are at the end of the projected endpoint with how they are at the end of the prolonged one. If Dale's final fricative had the same duration as Bethan's (which is much closer to the average duration for phrase final [s] reported by Crystal and House (1988)) his talk would have come to an end approximately 60% of the way into Bethan's final vowel ("interface"). This projected endpoint is indicated by the two-way arrow in Figure 3.

Dale's prolonging of his final fricative in line 7 of (3) has two outcomes. First, it allows him to complete a shift in gaze direction while his talk is in progress. Throughout the excerpt and until partway through line 12 Dale has his head oriented to the end of the room where Alex is standing next to the large screen. Dale's eyes (though not his head) begin to move to Bethan at roughly the same time as he begins his response in line 7. His eyes are directed across the table at Bethan when he starts his final fricative in "interface". At the projected endpoint of his talk Dale has started to move his gaze to his left, away from Bethan and towards the end of the room where Alex is standing next to the large screen. Dale's gaze direction at the projected endpoint is shown in the second screenshot accompanying line 7. By prolonging the final fricative in "interface", Dale gives himself time to direct his gaze to Alex before his talk ends. Note that Bethan does not have to modify her talk to allow for a coterminous shift in her gaze: she is gazing at her intended recipient (Alex) well before the end of her response, as she

starts [f] of “interface”. This chimes with the finding that speakers use and modify their talk to facilitate certain kinds of gaze behaviour (C. Goodwin, 1979, 1981). The second outcome for Dale of prolonging his final fricative is that he ends his TCU in the clear, i.e. out of overlap (Schegloff, 2000). This is in contrast to how things are at the projected endpoint which comes partway through Bethan’s final syllable.

By prolonging his final fricative in line 7 of (3), Dale gets to complete a change in gaze direction, and gets to conclude his talk in the clear. Why are these things significant? Speakers tend to gaze at recipients (C. Goodwin, 1979; Lerner, 2003; see also Rossano, 2012, on gaze at an addressee as a resource to elicit a response). Shifting his gaze to Alex before he finishes talking is a visible sign that the principal recipient of his talk is Alex, not Bethan. He is thus providing Alex with a solution to his repair initiation, rather than (for instance) offering his talk to Bethan for ratification. By continuing his talk beyond Bethan’s, Dale ensures that when Alex responds (as he surely will to work on the responses to his repair) Alex’s response will be closer to Dale’s than it will be to Chris’s or Bethan’s: Dale will therefore be the *de facto* provider of the solution.

In summary, at the end of his simultaneous response in (3) Dale produces an unexpectedly long fricative. Fricatives with the duration of Dale’s production at the end of line 7 (490 ms) are rare: Crystal and House (1988) give as 49 ms as the standard deviation for phrase-final [s] which has an average duration of 130 ms. It is only by considering issues in the local management of interaction (gaze, recipiency and sequential organisation) that what might otherwise be considered an unusual production can be understood.

The acoustic divergences dealt with in (2) and (3) have been about matters of duration and timing. In the final example of acoustic divergence, shown in (4), there are divergences in duration, pitch features and phonation types. The example is taken from the fourth meeting of a team tasked with designing a remote control for a television (cf. (3) where a different design team is working on the same task). In the minutes leading up to (4) the team has been discussing financial aspects of their favoured design led by Anna who is the project manager. Anna is working with a spreadsheet on her laptop computer, recording the features of the design so that costs can be established. The other team members can see the spreadsheet on a large screen at one end of the room. The spreadsheet is shown in Figure 4, as it was at the end of the meeting.

===== FIGURE 4 ABOUT HERE

Following an initial calculation the production cost for their design was established as €13.70. The team has been talking about how costs can be reduced to meet their target production price of €12.50, and have changed certain design features in an attempt to reach that target. Barbara has queried an increase in the overall cost displayed on the screen just prior to the excerpt transcribed in (4) .

(4) ES2008d-1533

First screenshot, clockwise from top left: Charles, Barbara, Anna, Diane. Out of shot to the right is a large screen showing what is on the screen of Anna's laptop computer.

▽

1 Bar: it was thirteen and now it's fifteen
2 (1.4)
3 Ann: no okay maybe not
4 (0.6)
5 Ann: I don't know what just happened
6 (1.0)

▽

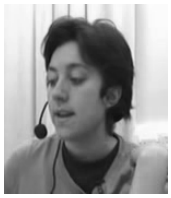
7 Ann: now [it's tw]elve mhhh
8 Dia: [(we're)]
9 Bar: oh [awr:
10 Cha: [what was our (.) target price again

▽

▽

11 Ann: .thh[hh **twelve point five**] mh[h[h h h h h h] h h]
12 Bar: [**twelve point five**] [h[eh hey]
13 Cha: [twelve point five] so]we're [just
14 Ann: [.hhhh

△ △



15 Ann: so w[e're okay
16 Cha: [just about there
17 Bar: we're [all set then]
18 Ann: [I thinkhh]
19 (1.0)
20 Ann: ye[s::::::]::
21 Dia: [( )]
22 (1.6)
23 Bar: ish hih

During Barbara's turn at line 1 Anna is operating the trackpad on her computer, presumably in an attempt to investigate the unexpected result in the spreadsheet. Anna refers to unexpected behaviour ("I don't know what just happened", line 5) before announcing a new price ("now it's twelve", line 7). This new price is receipted by Barbara (line 9). Charles seeks confirmation of their target price ("what was our (.) target price again", line 10). Anna

and Barbara provide the simultaneous response “twelve point five”, i.e. €12.50 (lines 11-12). This simultaneous response is receipted by Charles’s repetition of it (“twelve point five”, line 13).

Figure 5 provides visual representations of the simultaneous responses in (4). These three representations (top to bottom: spectrogram, pitch trace, waveform) are shown in horizontal alignment with each other. Orthographic labels and boundaries are shown along the top of the figures. The pitch traces show the relative position of each response in its speaker’s range. Even though the differences are readily apparent from listening, to measure and quantify the placement of talk in a meaningful way the speaker’s pitch range and other factors concerning the perception of pitch need to be taken into account (Walker, 2013). Zero represents the bottom of the speaker’s normal speaking range (= baseline), the top of each trace indicates the top, and the dotted line above zero indicates the speaker’s median pitch as an estimate of the middle of the speaker’s normal speaking range. All measures were established on the basis of corrected pitch traces of one minute of representative speech from each participant; this method for estimating a speaker’s range is used throughout this paper. The traces are plotted on a nonlinear semitone (ST) scale: an increase of 12 ST (an octave) corresponds to a doubling in frequency.

===== FIGURE 5 ABOUT HERE

The following are acoustic differences between the two versions:

1. There is a difference in duration: from the release of the initial plosive to the end measures 670 ms for Anna and 820 ms for Barbara. This difference is mostly due to the duration of the final sound which is much longer in Barbara’s response (180 ms for Barbara, 30 ms for Anna).

2. Both responses have a main accent on the final syllable (“five”). However, the final syllables have different pitch contours: Anna’s response ends with very low falling pitch whereas Barbara’s ends with a rising-falling-rising contour. Overall the placement of the utterance in the speakers’ ranges is different: excluding the very low pitched final syllable which ends far below the bottom of her normal speaking range, the mean pitch of Anna’s response is 2.7 ST above her baseline. The mean pitch of Barbara’s response is 4.8 ST above her baseline pitch and ends 0.7 ST above her median pitch. The pitch span (the range of pitch values the utterance covers) for the two responses is different: Anna’s response has a span of 2.3 ST (again, excluding the very low pitched final syllable); the pitch span of Barbara’s response is 3.9 ST.

3. The vowel portion of Anna’s final syllable (“five”, starting at approximately 0.5 s in Figure 5a) exhibits creak phonation throughout. Characteristic irregular low-frequency vibrations of the vocal folds are reflected in the irregular peaks in the waveform and irregular striations in the spectrogram from the onset of the final vowel. In Barbara’s version there is creaky voice for 30 ms (starting 140 ms into the final vowel at approximately 0.6 s in Figure 5b, where the computer software has been unable to make appropriate pitch estimations).

In summary: Barbara’s response is longer with a longer final sound; Anna’s response ends with very low pitch whereas Barbara’s ends higher with a final rising-falling-rising contour; and Anna’s response ends with creak through the whole of the final vowel whereas Barbara’s only has a short period of creaky voice towards the end of the final vowel.

How can the acoustic divergences between the simultaneous responses in (4) be accounted for? They are connected to the different stances being taken by Anna and Barbara to what is being said. Anna's final syllable has low pitch and creak phonation throughout the vowel portion. This is in line with her tendency to produce creak at the end of TCUs with low pitch. Creak and low pitch are evident at the end of Anna's TCUs in line 3 and line 7. A spectrogram, waveform and pitch trace of Anna's TCU in line 7 is shown in Figure 6.

===== FIGURE 6 ABOUT HERE

Creak phonation begins at 0.39 s in the figure, partway through the final syllable. This is reflected in the irregularly spaced vertical striations on the spectrogram and a discontinuity (sudden drop) in the pitch trace. Ogden, Hakulinen, and Tainio (2004) argue that phonetic design can be used as a means for marking out what is being said as conveying 'no news'. This is what Anna seems to be doing here: in ending her TCU in line 11 with creak and low pitch Anna is ending her response in what is, for her, a familiar and typical way. In doing so she is marking out what she is saying as 'nothing special'.

The production of "twelve point five" by Barbara at line 12 is different from Anna's at line 11 with rising-falling-rising pitch ending higher in her range. A number of researchers have documented connections between pitch height and modulation and the speaker's stance to what is being said. In particular, it has been argued that an 'emphatic' style can be used to signal heightened emotive involvement (Selting, 1994; see also Couper-Kuhlen 2012, M. H. Goodwin, Cekaite, & Goodwin, 2012, M. H. Goodwin, Goodwin, & Yaeger-Dror, 2002, Goudbeek & Scherer, 2009, Ogden, 2006). A cautious account of the pitch features of Barbara's TCU treats the pitch features as marking out the speaker's attitude to what is being said as something other than neutral. Wennerstrom and Siegel (2003) identified 135 points of possible syntactic completion followed by turn shift (cf. Barbara's TCU at line 12 which ends at a point of possible syntactic completion and is followed by a turn from Charles). The majority of those points (110 = 82%) were categorised as having a high rise or low final intonation boundary, neither of which apply here. At the very least, the pitch features of Barbara's response at line 12 of (4) set it apart from normative expectations, and as 'out of the ordinary'.

What is said in the simultaneous response in (4) is not simply an announcement of a target price: all of the participants (including those who make the announcement, i.e. Anna and Barbara) can compare the figure given in that announcement with the team's current position and reach a conclusion as to whether the team is where it should be in terms of its spending. In other words each of the participants can display a stance to what has been said. The different stances being taken by the participants becomes clear in the development of the sequence after the simultaneous response where stance continues to be negotiated at least until Barbara's addition of "ish" at line 23. Just after Barbara produces "twelve point five" (line 12) she produces the assessment "hey hey" (line 4) with rising-falling pitch and total pitch excursion of some 12.8 ST. (From the sequential organisation of the talk up to its point of occurrence it may seem that Barbara's "hey hey" could be responsive to having produced the same talk at the same time as Anna, cf. the mention of jinxing above. However, neither Barbara - who does not make eye contact with Anna, although Anna has her gaze directed at Barbara - nor any of the other participants treat it this way, whereas they do go on to produce assessments of the team's position.) Shortly after, when Anna assesses the situation she gives the reserved "so we're okay" (line 15), which is produced with a total pitch excursion

(excluding the final creaky portion at the end of "okay") of only 4.5 ST. This difference in pitch span - greater for Barbara than for Anna - is part of how they are marking out the different stances they are taking (on phonetic features and the marking of stance in interaction see Couper-Kuhlen, 1986, 2012, Freese & Maynard 1998, Maynard & Freese, 2012, Ogden 2006). Experimental studies have shown a consistent relationship between 'joy/elation' and an increase in pitch span (Scherer, 1986, Goudbeek & Scherer, 2009), as well as showing that the pitch span for positive connoted statements is wider than for negative connoted ones (Reckling & Kügler, 2011). The development of the sequence following the simultaneous response thus provides strong evidence that the observed differences in pitch features of the simultaneous responses reflect different stances being taken by Anna and Barbara (more positive for Barbara than for Anna).

It was said above that the final sound in the simultaneous responses in (4) differ in their duration, with Anna's being much shorter than Barbara's. However, such a short production at the end of line 11 may not be so surprising given Anna's speaking norms: Anna's production of the same sound at the end of line 7 also measures just 30 ms. What is remarkable about the end of Anna's turn at line 11 is that she ends it with a prolonged labiodental (lower lip and upper teeth) closure. This closure is held from the end of the final sound of "five" (line 11), through the long nasal outbreath which follows her turn and up to the onset of her inbreath in line 14. By the end of the inbreath her lower lip is in an 'at rest' position. This is very different from the end of line 7, where immediately after her turn-final labiodental fricative Anna brought her lips together and breathed out through her nose. It has been shown that held articulations can project more talk (Local & Kelly, 1986; Local & Walker, 2005). However, those cases involve the holding of glottal rather than supraglottal closure and by definition cannot be accompanied by breathing. Whereas the pitch and phonatory features of Anna's response mark out what she is saying as 'nothing special', the prolonging of this labiodental closure marks out what she is saying as 'out of the ordinary'. This signal of 'out of the ordinary' is all the stronger given the position of her eyebrows during this turn. In her turn at line 7 ("now it's twelve") Anna raises her eyebrows as she produces the syllable which carries the focal accent ("twelve"). This is as we might expect given that there is usually an increase in eyebrow movements on focal accents (Granström & House, 2007). She raises her eyebrows much earlier in the production of the simultaneous response: her eyebrows begin to raise during her inbreath at the start of line 11, are fully raised by the end of the inbreath, and remain raised throughout her production of the simultaneous response. The production of this TCU with eyebrows raised throughout, rather than only on the focal accent, also marks out what is being said as 'out of the ordinary'. Rather than working to convey the same message, visible and audible features of Anna's turn at line 11 are each signalling something different: 'nothing special' in the case of certain audible characteristics (pitch and phonation type), and 'out of the ordinary' in the case of certain visible ones (prolonged labiodental closure and raised eyebrows).

By examining the development of the sequence it becomes clear that Barbara's stance with regards to the implication of the announcement is that it is wholly positive, whereas Anna is more reserved. They begin to display these stances in part through the pitch characteristics of their simultaneous responses. Anna and Barbara also convey something of the different stances they are taking through the duration and nature of the final sound at the end of the simultaneous responses ("twelve point five"), and their facial expressions.

Using (2)-(4) as exemplars, this section has shown that there can be acoustic divergences between simultaneous responses in interaction, where those responses are identical (2, 4) or

near-identical (3) in their lexico-syntactic format. It has been shown that acoustic divergences can be accounted for and understood by considering the local interactional tasks the participants are managing as they produce their response. It would weaken the case that acoustic divergences can arise from differences in the interactional work being done if simultaneous responses are never the same in pronunciational terms. Conversely, being able to show that simultaneous responses can exhibit acoustic equivalence would strengthen the claim that acoustic divergence does not arise from inevitable variation in speech, and therefore can be harnessed to handle different kinds of interactional work. The final case to be considered here is an example of acoustic equivalence.

4 Acoustic equivalence in simultaneous responses

A case of acoustic equivalence is shown in (5). The participants are the same as in (2), again discussing movies to be screened at a movie club. Chris has made various suggestions for the screening including *Silence of the Lambs*. David has asked for clarification of the plot of the movie prior to the transcribed excerpt. Clarification has been provided by the three other participants, mostly Adele and Ben.

(5) IB4010-1271
Left to right: David, Ben, Adele, Chris.

```
 1  Ben:   so on the picture you can see uh (.) Jodie Foster
 2         [and] uh (.) .pt (.) [ha hah]
 3  Dav:   [yeh]                [.pt.hh]
 4  Ben:   hmh(h)h [the
 5  Dav:           [the moth
 6  Dav:   who is playing the [ps uh- pseek]opat again
 7  Ben:                     [ the  Moth  ]
 8              (0.5)
 9  Dav:   [(you don't)]
10  Ade:   [  tuuh     ] which (.) the one who's in jail [   o]r
11  Ben:                                                  [.ptk]
12  Ade:   th[e one who-]
13  Dav:     [  yeah    ]
14              (0.2)
```

▽

```
15  Ade:   [Anthony Hopk[ i  n  [s]
16  Ben:   [Anthony Hopk[ i  n  [s]
17  Chr:                [Anthon[y]
18  Dave:                      [o]kay yeah
19              (0.3)
20  Ade:   the other one:: I don't remember
```

At line 6 David asks about one of the actors in the movie ("who is playing the pseekopat [psychopath] again"). There is a request for clarification from Adele as to which character he is referring to ("which (.) the one who's in jail or the one who-", lines 10-12). Following David's response to the first part of the request for clarification ("yeah", line 13) Adele and Ben both respond with conditionally relevant responses when they say "Anthony Hopkins" (lines 15-16).

The responses in (5) are striking in their acoustic similarity, notwithstanding differences arising from the gender and language background of the participants (Adele is a female native speaker of English; Ben is a male native speaker of Romanian). Figure 7 shows waveforms of the responses.

===== FIGURE 7 ABOUT HERE

The response from Adele starts approximately 20 ms after Ben's and therefore for all practical purposes at the same time. The overall duration of each response is very similar: Adele's response measures 1.11 s while Ben's measures 1.08 s. The participants are thus speaking at a very similar overall rate. The lag between acoustic landmarks in the responses is never greater than 50 ms (mean = 24 ms) and Adele's acoustic landmarks are after Ben's in all but one case (landmark *b*, which is synchronised). Adele and Ben are both speaking at the same rate as each other throughout the utterances, without significant differences in relative speed.

The simultaneous responses in (5) are equivalent in their overall pitch contour, the pitch span and the placement of the utterance in each speaker's range. Figure 8 presents pitch traces of the simultaneous responses in (5). The traces have been normalised for duration: the same syllable in each trace takes up the same amount of horizontal space; the amount of horizontal space taken up by each syllable reflects the average proportion of the utterance taken up by that syllable.

===== FIGURE 8 ABOUT HERE

The overall pitch contour of the two responses is the same, with falling pitch over the first part of the utterance as far as the main falling accent on "Hopkins", followed by rising pitch on the final syllable. The two responses cover a similar proportion of each speaker's range. The placement of the utterances in each speaker's range is similar: the mean for the utterances are within 2 ST of the relevant speaker's baseline pitch.

The two simultaneous responses are equivalent in terms of their loudness, with both responses in line with each participant's normal speaking volume. Measurements of peak intensity of each syllable in the simultaneous response and of syllables in each participant's preceding talk in (5) provides corroborative acoustic evidence for loudness equivalence. The average of the peak intensities for Ben's "so on the picture (you) can see Jodie Foster" (line 1) is 83.5 dB; the average for his contribution to the simultaneous response is slightly lower (mostly due to a relatively quiet final syllable) at 78.8 dB. The average of the peak intensities for Adele's "which one the one who's in jail or the one who" (line 10-12) is 73.9 dB; the average for her contribution to the simultaneous response is 74.3 dB. As well as being in line with each participant's normal speaking volume, neither response shows the sorts of increases in loudness found in turn-competitive overlap (French & Local, 1983), but rather have even loudness until a quieter final syllable.

In summary, the example in (5) shows that it is possible for two simultaneous responses to be acoustically equivalent. The acoustic equivalence of the two responses in (5) throws into relief simultaneous responses like those in (2)-(4) where there are noticeable differences between the responses. This strengthens the case that where there are acoustics divergences these must not be discarded as a consequence of inevitable variation in speech. Rather, they must be considered for their possible connections with the organisation of interaction (cf. Heritage, 1989, Local & Walker, 2005).

5 Summary and implications

This paper serves as a demonstration of how the analytic techniques developed within a particular research paradigm can be applied to the details of strips of talk-in-interaction. That paradigm combines analysis of phonetic details with analysis of the sequential organisation of the talk following the principles of CA. The level of detail in the analysis is not a matter of analytic indulgence (Local & Walker, 2005), but a necessary step in understanding the communicative function of phonetic details. This paper has demonstrated the need to attend to the organisation of interaction in a principled and systematic fashion when trying to account for the different pronunciational shapes of utterances. It has reported on sequential and phonetic aspects of simultaneously produced lexically equivalent (and often lexically identical) responses.

It has been shown that simultaneous responses may exhibit acoustic divergences (2-4), or may be acoustically equivalent (5). Potentially any phonetic parameter can be a source of acoustic divergence: cases of acoustic divergence dealt with here have included timing and duration, articulation, phonation and pitch. It has been shown that acoustic divergences can arise from the interactional contingencies conversational participants are dealing with: the pronunciational shapes of utterances are shaped by the interaction. These contingencies include differences in knowledge (2), issues in selecting a recipient for the response (3), and differences in evaluative and epistemic stance (4). These findings echo those of other research into the phonetics of talk-in-interaction (see Walker, 2013), and carry an important implication for research on phonetic variation in naturalistic talk: that any such research which disregards the local organisation of interaction will always be incomplete.

Acknowledgements

Thanks to John Local and Traci Walker, as well as Paul Drew and the reviewers, for comments which helped me refine and develop the paper in various ways.

References

Carletta, J. (2007). Unleashing the killer corpus: Experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation, 41* (2), 181-190.

Clayman, S. E. (2013). Turn-constructional units and the transition-relevance place. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 150-166). Oxford: Wiley-Blackwell.

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

Crystal, T. H. & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, 83* (4), 1553-1573.

Couper-Kuhlen, E. (1986). *An introduction to English prosody.* London: Edward Arnold.

Cummins, F. (2002). On synchronous speech. Acoustics Research Letters Online, 3 (1), 7-11. Retrieved from http://scitation.aip.org/content/asa/journal/arlo

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics, 31* (2), 139-148.

Cummins, F. (2006). Probing the dynamics of speech production. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in empirical prosody research* (pp. 211-228). Berlin: Walter de Gruyter.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics, 37* (1), 16-28.

Cummins, F., Li, C., & Wang, B. (2013). Coupling among speakers during synchronous speaking in English and Mandarin. *Journal of Phonetics, 41* (6), 432-441.

French, P. & Local, J. (1983). Turn competitive incomings. *Journal of Pragmatics, 7* (1), 17-38.

Freese, J. & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society, 27* (2), 195–219.

Fry, D. B. (1975). Simple reaction-times to speech and non-speech stimuli. *Cortex, 11* (4), 355-360.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97-121). New York: Irvington.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Goodwin, M. H., Cekaite, A., & Goodwin, C. (2012). Emotion as stance. In M. Sorjonen & A. Peräkylä (Eds.), *Emotion in interaction* (pp. 16–41). Oxford: Oxford University Press.

Goodwin, C. & Goodwin, M. H. (1987). Concurrent operations on talk: Notes on the interactive organization of assessments. *IPrA Papers in Pragmatics, 1* (1), 1-54.

Goodwin, M. H., Goodwin, C., & Yaeger-Dror, M. (2002). Multi-modality in girls' game disputes. *Journal of Pragmatics, 34* (10-11), 1621-1649.

Goudbeek, M. & Scherer, K. R. (2009). Expression of emotion - Vocal expression. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (pp. 404–406). Oxford: Oxford University Press.

Granström, B. & House, D. (2007). Inside out - acoustic and visual aspects of verbal and non-verbal communication. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the ICPhS XVI* (pp. 11-18). Saarbrücken.

Hay, J. & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology, 36* (89-103).

Hayashi, M. (2013). Turn allocation and turn sharing. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 167-190). Oxford: Wiley-Blackwell.

Heritage, J. (1989). Current developments in conversation analysis. In D. Roger & P. Bull (Eds.), *Conversation: an interdisciplinary perspective* (pp. 21-47). Clevedon: Multilingual Matters.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies, 9* (2-3), 153-183.

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.

Lerner, G. H. (1993). Collectivities in action: Establishing the relevance of conjoined participation in conversation. *Text, 13* (2), 213-245.

Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society, 32* (2), 177-201.

Lerner, G. H. (2004). Collaborative turn construction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 225-256). Amsterdam: John Benjamins.

Local, J. (2005). On the interactional and phonetic design of collaborative completions. In W. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 263-282). Mahwah, NJ: Lawrence Erlbaum.

Local, J. & Kelly, J. (1986). Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies*, 9, 185-204.

Local, J. & Walker, G. (2005). Methodological imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech. *Phonetica, 62* (2-4), 120-130.

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26* (11), 2530-2539.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4* (1-3), 55-73.

Maynard, D. W. & Freese, J. (2012). Good news, bad news, and affect: Practical temporal and ''emotion work'' in everyday life. In M. Sorjonen & A. Peräkylä (Eds.), *Emotion in interaction* (pp. 92–112). Oxford: Oxford University Press.

Moore, J., Kronenthal, M., & Ashby, S. (2005). *Guidelines for AMI speech transcriptions, Version 1.2.* Retrieved from http://groups.inf.ed.ac.uk/ami/corpus/Guidelines/speech-transcription-manual.v1.2.pdf

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 'no news' with stylization in Finnish. In E. Couper-Kuhlen & C. E. Ford (Eds.), *Sound patterns in interaction* (pp. 299-334). Amsterdam: John Benjamins.

Reckling, F. & Kügler, F. (2011). Pitch range in positive and negative connoted statements of German. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the ICPhS XVII* (pp. 1670-1673). Hong Kong.

Rossano, F. (2012). *Gaze behavior in face-to-face interaction* (Doctoral dissertation, Radboud University Nijmegen).

Roud, S. (2010). *The lore of the playground: One hundred years of children's games, rhymes & traditions.* London: Random House.

Sacks, H. (1992). *Lectures on conversation*. Oxford: Blackwell.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70* (6), 1075-1095.

Schegloff, E. A. (1995). Parties and talking together: Two ways numbers are significant for talk-in-interaction. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in social organization and embodied activities* (pp. 31-42). Washington, D.C.: University Press of America.

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

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychology Bulletin, 99* (2), 143-165.

Selting, M. (1994). Emphatic speech style - with special focus on the prosodic signaling of heightened emotive involvement in conversation. *Journal of Pragmatics, 22* (3-4), 375-408.

Sidnell, J. & Stivers, T. (Eds.). (2013). *The handbook of conversation analysis.* Oxford: Wiley-Blackwell.

Walker, G. (2012). Coordination and interpretation of vocal and visible resources: 'Trail-off' conjunctions. *Language and Speech, 55* (1), 141-163.

Walker, G. (2013). Phonetics and prosody in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 455-474). Oxford: Wiley-Blackwell.

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

Wennerstrom, A. & Siegel, A. F. (2003). Keeping the floor in multiparty conversations: Intonation, syntax, and pause. *Discourse Processes, 36* (2), 77-107.

Figure 1: Waveforms of the simultaneous responses in (2), lines 12-13 showing the relative timing of acoustic landmarks in the two responses.

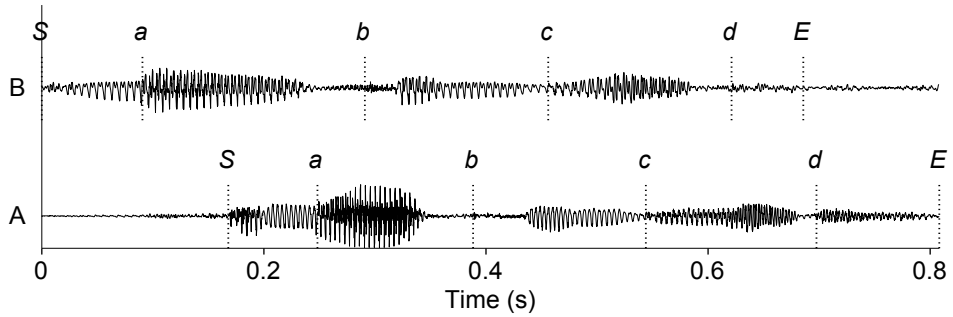


Figure 2: Slide shown on large screen for the duration of (3).

Closing

- Next meeting starts in 30 minutes
- Individual actions
  - ID: the working design
  - UID: the technical functions design
  - ME: the user requirements specification
- Specific instructions will be send to you by your personal coach.

Figure 3: Waveforms of the simultaneous responses in (3), lines 6-7 showing the relative timing of acoustic landmarks in the two responses. The two-way arrow indicates where Dale's talk would have ended relative to Bethan's if his final fricative ("interface") had the same duration as hers. The label *S/a* in Dale's waveform indicates that the start (*S*) and landmark *a* (the start of the first vowel in "interface") are one and the same.

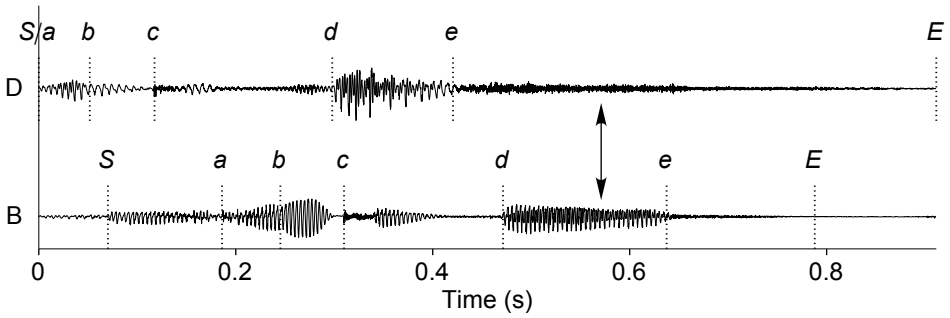


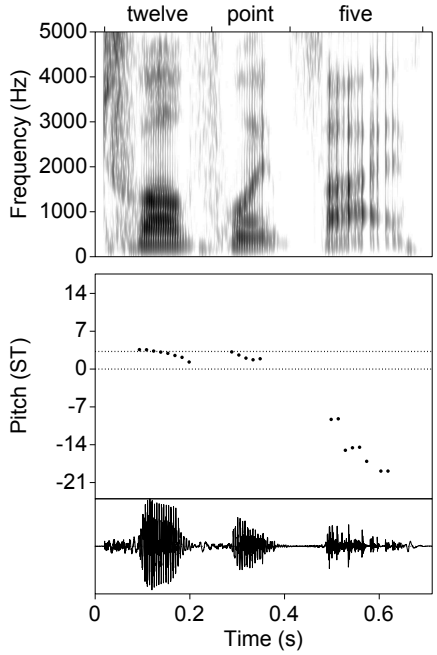
Figure 4: The spreadsheet Anna is working with during (4) showing the costs of the prototype design.

	Price	#	Sum
Energy source			
hand dynamo	1		0
battery	2		0
kinetic	3	1	3
solar cells	4		0
Electronics			0
simple chip on print	1	1	1
regular chip on print	2		0
advanced chip on print	3		0
sample sensor / sample speaker	4		0
Case			0
uncurved/flat	1		0
single curved	2	1	2
double curved	3		0
case material supplements			
plastic	0		0
wood	1		0
rubber	2	1	2
titanium	3		0
special colour	0.5	1	0.5
Interface			0
push button	0.5	6	3
scroll wheel	1.3		0
integrated scroll wheel/ pushbutton	2.5		0
LC display	3		0
button supplements			
special colour	0.2	1	0.2
special form	0.4		0
special material (rubber, wood, titanium)	0.6	1	0.6
Total			12.3

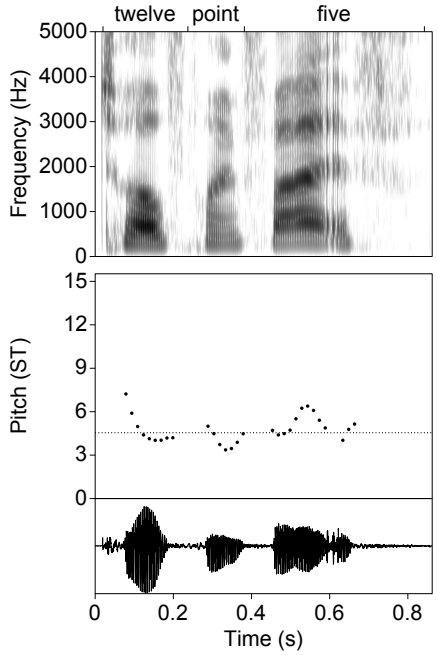
Fill in the number of components you plan to use in your device and the total cost will be automatically calculated.

Figure 5: Spectrograms, pitch traces and waveforms of the simultaneous responses in (4). Each subfigure starts 50 ms before the release of the plosive at the start of "twelve" (the first

boundary along the top of each subfigure) and ends 50 ms after the end of the final sound in "five" (the last boundary in each subfigure).



(a) Anna, line 11; the scale has been extended two octaves below her normal speaking range in modal voiced phonation to show frequencies during portions with creak phonation



(b) Barbara, line 12

Figure 6: Spectrogram, pitch trace and waveform of Anna's TCU in line 7 of (4).

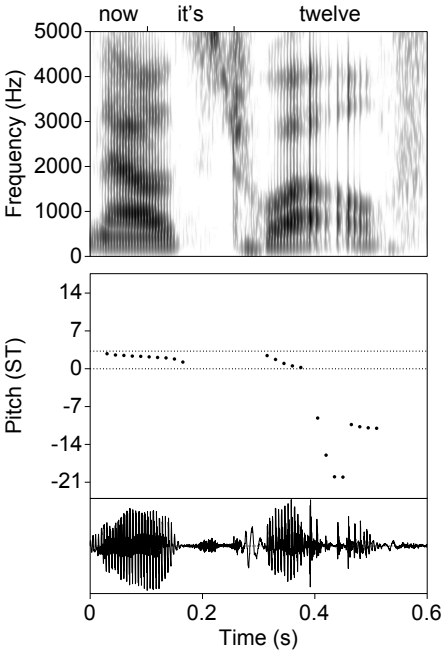


Figure 7: Waveforms of the simultaneous responses in (5), lines 15-16 showing the relative timing of acoustic landmarks in the two responses

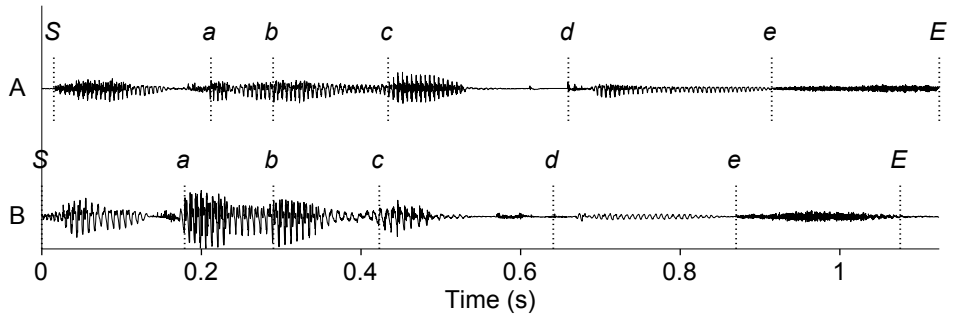


Figure 8: Pitch traces of the simultaneous responses in (5).