Turn-sharing revisited:

An exploration of simultaneous speech in interactions between couples

Stefan Pfänder, University of Freiburg, Germany

Elizabeth Couper-Kuhlen, University of Helsinki, Finland

Abstract

This paper investigates one particular type of simultaneous speech, namely turn-sharing, in the Freiburg Sofa Talks, a corpus of video-recorded dyadic conversations between partners, friends, and siblings who are recollecting events they have experienced together in the past. The focus is on interactions in German and French. In turn-sharing, participants aim at saying the same thing at the same time, using these moments to convey something to each other, and occasionally to a third party in the room. We identify two different types of turn-sharing, choral performance and chiming in, which are brought off by different micro-practices with verbal, prosodic, and bodily resources. Each type achieves something different interactionally, either displaying a shared affective stance towards something in an alternative world or embodying an epistemic claim to know as much as the main speaker. We conclude that choral performance and chiming in are two sedimented formats for turn-sharing that are achieved with different practices using semiotic resources that are comparable, if not identical, across languages.

Keywords

Turn sharing, simultaneous speech, affective stance, epistemic claim, embodied practices

Introduction[1]

This paper investigates simultaneous speech in the Freiburg Sofa Talks, a corpus of video-recorded dyadic conversations between partners, friends, and siblings who are recollecting events they have experienced together in the past. While the corpus includes interactions conducted in French, German, Italian, and Spanish, this paper reports exclusively on simultaneous speech in the French and German interactions. It examines one particular type

---

of simultaneous speech, namely turn-sharing, occasions when participants aim designedly at saying the same thing at the same time. On these occasions they use simultaneous speech to convey something to each other and to a third party in the room. We explore two issues with respect to such occasions: (a) how the participants manage to say the same thing at the same time, and (b) what they achieve in doing so. The analysis considers not only verbal but also embodied dimensions of simultaneous speech.

1. Types of simultaneous speech

Various strands of research have touched on different types of simultaneous speech, moments in time when two (or more) participants speak at once in talk-in-interaction.

Joint speech, for instance, has long been the focus of scholars working on talk in group interactions. The latter include established institutionalized interactions, such as choral prayer in church and chanting in more or less improvised (or grassroot) rituals, e.g., political marches or demonstrations (for a detailed investigation of convergences and divergences between these types of joint speech, see Cummins 1997, 2001, 2014, 2018). It is worth noting that in these scenarios hand gestures are often combined with chanted speech. Whereas in such cases both bodily gestures and language play a role, in other less ritualized contexts language may play a minor role, for example, in the case of shared *audience laughter* or *group applause*, which is typically prompted by trained speakers or comedians (Atkinson 1984; Heritage & Greatbatch 1986; Heritage & Clayman 2010, ch. 18).

Yet another type of context in which joint speech is the default option and not the exception are the *greetings and farewells* of everyday life. When we enter a room filled with a crowd of people, our co-interlocutors are more likely to answer our greeting chorally than sequentially (Pillet-Shore 2012). And in telephone closings, simultaneous talk is the rule rather than the exception (Auer 1990). Certainly, greetings and farewells have something ritualistic about them, but they are part of everyday talk-in-interaction as well.

In multi-party conversations, simultaneous speech can emerge when two or more participants start up a turn at the same time: this is known as a 'simultaneous start' (Sacks et al. 1974) or a 'simultaneous response' (Walker 2016). Overlapping stretches of talk may, however, also be due to the early entry of a self-selecting next speaker. At least three recurrent scenarios for early entries have been suggested. It has been found that the next speaker may enter a turn before its ending when the end is projectable (see Jefferson 1973 on

precision timing). In this case, the early entry helps to ensure the right to speak for the self-selecting participant. Then again, early entries can also occur by accident when self-selecting next speakers identify a possible place of transition which, however, was not intended to be the end of the previous speaker's turn (see Jefferson 1973; Drew 2009). Not all early entries are cooperative in nature: there are also competitive (French & Local 1983) and subversive incomings (Bolden et al., in press).

In some cases of simultaneous speech, two or more interlocutors not only share the turn, i.e., speak at the same time, but also utter the exact same words (Tannen 1989; Lerner 2002; Schwitalla 1993): see §2 below. It is this last type of simultaneity, choral speech, that is the focus of this article. Our aim is to contribute to a better understanding of how and why co-interlocutors manage to say the same thing at the same time.

2.   Prior research: Tannen, Lerner and beyond

There has been a handful of prior studies dealing with choral speech. The earliest of these was arguably Tannen (1989), who identified the phenomenon of *shadowing*, the repetition of what one speaker is saying to another delayed by a split-second. Here is one of her examples:

(1) Shadowing (Tannen 1989:92)
```
    4     DEBORAH     or Westerners tend to uh: ...
    5                 think of the bódy and the sóul
    6                 as two different things,
    7     CHAD                                  ⌊Right.
    8     DEBORAH     because there's no word
    9                 that expresses body and soul together.
   10→   CHAD                              ⌊Body and soul together.
   11                 Right.
```

What Tannen calls shadowing occurs in line 10: here, the recipient repeats what he hears as he hears it with only a split-second delay (p. 89). Yet in contrast to Tannen's description, which treats shadowing as an "automatic" form of "repetition", we will instead argue that such simultaneous speech is neither 'repetition' nor 'automatic'. Rather, it involves two (or more) participants saying the same thing at the same time by design.

This is the position that Lerner (2002) takes; indeed, our study is much indebted to his seminal work. Lerner describes **choral speech** as a form of participation that involves participants who aim to co-produce all or part of a turn-constructional unit (TCU) in unison

with one another, designedly matching words, voicing, and tempo (p. 226). Here is one of his examples (the participants are reminiscing about a line from a well-known movie featuring the Three Stooges):

(2) Choral speech (Lerner 2002: 232)
```
14     MIC:   Remember that?=
15     SHA:   =Yeah
16→    MIC:   [I kne:w you were co:ming so I ba:ked a ca:ke.
17→    SHA:   [You knew I you coming, so I baked a cake
18→    VIV:   [I knew you were coming so I () ca(h)ke heheheh
```

As evidence for the designedness of this co-production, Lerner notes that Shane quickly shifts from 'I' to 'you' in line 17 in order to coordinate his rendition of the line with Michael's version (line 16). Moreover, as is visible in the video, Michael accompanies line 16 with a 'cake presentation' gesture and proceeds to 'conduct' the shared performance by superimposing gestural beats on 'baked' and 'cake' (2002: 233).

Choral speech as described by Lerner has the following characteristics:

o The joining in of recipients to an emerging turn at talk can be done *without* elicitation (or prompting) by the current speaker or *with* elicitation by the current speaker (p. 228).

o In cases *without* elicitation, it is typically the terminal item of a TCU that is co-produced. The heightened projectability of the trajectory of the turn, as well as its position in a sequence of actions, is what allows recipients to anticipate what will come next and to co-produce it.

o In cases *with* elicitation, one technique involves initiating shared reminiscence through a "reminiscence recognition solicit" such as '*Remember that?*' (see line 14 of (2) above) (p. 232).

o Choral co-production can be used for a number of effects. One of these is to accomplish an action in a conjoined fashion for a third participant. Choral co-production can also serve as a way for the addressed recipient to demonstrate agreement with what is being said (p. 237). In the case of mutual reminiscence, it can be a way to "produce an action as a reciprocal reminiscence of a known- (or imagined- or created-)-in-common experience" (p. 239). This is presumably what happens in (2) above. It allows participants to "establish or sustain their

entitlement to co-authorship/ownership of experience" and to "demonstrate their appreciation of their co-participant's shared entitlement" (p. 239).

In what follows, we will be building on Lerner's analysis of choral speech, while at the same time extending and deepening it – particularly with reference to the microanalytic means which participants deploy to coordinate speech simultaneously. In the process of asking how simultaneous speech is brought off, we have discovered at least two different ways in which participants can produce the same words at the same time. These map to some extent onto Lerner's 'elicited' and 'unelicited' types of choral production but, as we show, they are accomplished through different practices and correlate with different achieved effects. Although both types would presumably be covered by Lerner's notion of "choral speech", we prefer to distinguish them terminologically: in one case, we speak of 'choral performance' and, in the other, of 'chiming in'. Due to the specifics of our data collection (see §3 below), both involve what Lerner (2002) would call "mutual reminiscence".

3.   Data and methodological procedure

The Freiburg Sofa Talks Corpus comprises 208 video recordings ranging in duration from 10 to 40 minutes. In each recording, two people collaboratively reconstruct their shared experiences while sitting on a sofa together. The corpus contains material from four European languages (Italian, Spanish, French and German).[2] For our study, we concentrate on French and German data. The protagonists in each of the recordings have known each other for quite some time: they may be close friends, siblings or married couples. They were instructed to jointly recall things they have experienced together in the past. Shared experience of the narrated events in question was critical for inclusion in the corpus. Before each recording starts, the two participants are explicitly asked to tell their stories together, which is intended to guarantee equal epistemic authority as well as an equal right to speak. The two protagonists are free to choose the episodes they will talk about beforehand. The recording takes place in the presence of a third person -- a close friend, another sibling, or a neighbor – who, however, does not actively intervene in the reminiscing. This has two implications. First, the two participants on the sofa self-manage the process of choosing their topics and of assigning

---

[2] More languages will be included in the following years. The corpus is still under construction and not yet open access, but the first author of this article can be consulted for more information: http://www.romanistik.uni-freiburg.de/pfaender/769.

speaker roles. Second, the recipient is not exclusively the partner on the sofa, since the third person, who is sitting next to a fixed camera, may constitute a further possible addressee.

To be sure, one can ask, are these naturalistic or experimental data? The design of the recordings was chosen in order to approximate naturally occurring conversation as much as possible, while still being suitable for comparison as in an experimental setting. On the one hand, it is a recognizable social routine for partners, close friends, or siblings to sit together and reminisce about things they have experienced. This closeness to actually occurring social practices may shed light on why we find many of the interactional practices described for more authentic everyday interaction in our data: stories emerge from prior talk, the right to speak as well as epistemic and affective stances are often negotiated; there are numerous instances of interruption and overlapping speech, etc. On the other hand, being asked to self-organize turn-taking and topic choices while sitting next to one another in front of a camera and a known but silent third person is not part of anyone's daily routine. However, we believe the advantages of this method of data collection outweigh its drawbacks.

For our study, we have made a collection of 32 instances of choral production, of which we discuss – by way of exemplification - four instances (two French and two German) in the following sections. To anticipate the argument made, our data show that simultaneous productions can be accomplished in at least two different ways, labelled here 'choral performance' and 'chiming in'. The claim based on our collection is that these variants of simultaneous speech not only come about through different ways, or practices, of mobilizing semiotic resources but that they achieve different purposes in interaction. Neither type of simultaneous speech is, however, accidental or coincidental. Rather, participants aim at coordinating their talk as a form of turn-sharing.

The remainder of this paper is organized as follows: in §4 and §5, we describe one prototypical case each of 'choral performance' and 'chiming in', first in German interaction and then in French. §6 explores the semiotic resources used to accomplish these two formats for simultaneous speech and §7 discusses what these formats allow participants to achieve interactionally. In §8, we consider language-specific aspects of bringing off simultaneous speech and in conclusion, discuss the implications of our observations for a multimodal conception of turn-sharing across languages.

4. The prototype of choral performance

A first prevalent type of simultaneous speech in our collection is similar to Lerner's 'I knew you were coming' example, cited in (2) above. We call it *choral performance* because (a) it comes about through the concerted efforts of the two participants to produce talk in unison and (b) its character is primarily performative: participants take pleasure in their co-production and at its closure display their mutual appreciation of each other and what they have produced together.

4.1 Choral performance in German interaction

We begin by examining a case of choral performance in German. The following data extract comes from an exchange between Anna and Emma, who are close friends and used to live in the same town but now live a couple of hours away from each other. Here they are reminiscing about a trip to Spain. As the breakfast was bad at their hotel, they started their day in a small café nearby. Anna now all of a sudden remembers the orange juice they had at the café:

(3) "Frisch gepressten Orangensaft" (‚Freshly squeezed orange juice')[3]

```
01 Anna:   und wir sind dann jeden morgen FRÜHstücken gegangen;
           and then we went out for breakfast every morning

02         aber das frühstück war TO::LL.
           but the breakfast was great

03 Emma:   [<<creaky>JA:>        ]
                      yeah

04 Anna:   [<<pp>richtig TOLL.>]
                   really great

05 Emma:    und vor Allem es war halt Echt schön
            and above all it was really nice

           dadurch dass es am STR[AND war]
           because it was on the beach

06 Anna:                         [JA::- ]
                                  yeah

07         <<all>das (hab'm wir ja)/-> (.)
                   that we really
```

---

[3] This and all following examples have been transcribed according to the GAT2-conventions (Couper-Kuhlen & Barth-Weingarten 2011)

```
08          OCH der oRANgensaft;=
            oh the orange juice

09          =<<dim>da hÄtt ich jetzt LUST drauf.>
                 I could go for that now

10          °hh:::

11→         [oh so_n <<rhyth, dim>frIsch geprEssten oRANgensa:ft->]
             oh like a              freshly squeezed orange juice

12→ Emma:   [ah ja (n)<<rhyth,dim>frIsch geprEssten oRANgensa:ft->]
             oh yeah (a)           freshly squeezed orange juice

13          (0.6)

14 Anna:    ((click))

15 Emma:    (j)a_[a:hh.      ]
            yeah   ohh

16 Anna:        [=allgEmein] hab ich grade HUNGer;
                 actually I'm hungry right now

17          <<all> aber eGAL->
                 but never mind

18          ((lacht))
            ((laughs))
```

As Anna suddenly recalls the orange juice at the Spanish café (line 08), she produces what Goffman would call a 'response cry' (Goffman 1978; Heritage 1984) and wishes aloud that she had some then and there (line 09). There follows an audible and lengthened inbreath by Anna (line 10) and then a highly rhythmic choral production[4] of *frIsch geprEssten oRANgensaft* 'freshly squeezed orange juice' said by Anna and Emma together (line 11 and 12). After a brief silence, Anna concludes the performance with an audible click[5] (line 14) and Emma confirms its closure with *ja* 'yeah' (line 15).

How do the participants manage to bring off this choral performance? On closer examination, it can be seen that they deploy a careful choreography that ensures near-perfect coordination. The simultaneous speech itself is preceded by three distinct preparatory phases, which allow the participants to cue and fine-tune their performance. In the first phase, the current speaker makes an abrupt gesture – here Anna's *OCH der oRANgensaft* (line 08), produced after she has prematurely broken off the prior unit.[6] This serves to put the second speaker on alert ('On your mark'). Second, the current speaker sustains this alert with a holding gesture – here the audible inbreath in line 10 ('Get set'). Third, a split-second

---

[4] The rhymicity of this choral production is described in more detail below.
[5] For more on clicks in everyday conversation see Wright 2011a+ b and Ogden 2013.
[6] For more on German *och*, see Golato (2012: 261-2).

transition period follows in which the second speaker joins in and the two make micro-adjustments to ensure coordination – here Emma's *ah jA[7] (n)* in overlap with Anna's *oh s_n* ('Go!'). The choral production *frIsch geprEssten oRANgensaft* is a culmination of these preparatory phases (lines 11 and 12).

The three phases are visualized in Figures 1-4 below. The symbol # in the transcript line refers to the exact moment of the screenshot above it. Fig. 1 shows the first phase: while Anna (on the left) utters *OCH der oRANgensaft,* she raises her head (visualized through motion tracking below by the ascending curve, which traces the movement of Anna's chin: see Graph 1). Fig. 2 shows the second phase: Emma's gaze is fixed on Anna, meaning she can both hear and see Anna's prolonged inbreath. Fig. 3 shows the fine-tuning phase and Fig. 4 the moment of the choral onset. Note that as the choral performance gets underway, Emma averts her gaze.

Fig. 1 Alert   **Fig. 2** Hold

08 ANN: OH der #oRANgensaft;=   10 ANN: #°hh:::

---

[7] Cf. Betz & Golato (2008: 59) on German *achJA* for expressing "relevant but just now recalled information".

**Fig. 3** Fine tune   **Fig. 4**   Start


```
11 ANN:[#h so_n <<rhyth, dim>.    #frIsch geprEssten oRANgensa:ft->]
12 EMM:[# ah ja (n) <<rhyth, dim> #frISCH geprEssten oRANgensa:ft->]
```

With the help of the Freiburg Motion Tracking tool VEO (Visualisation of Emergent Objects, cf. Alcón, Muz, Satti & Lahmann, submitted), we can give a more abstract account of the flow of movements in this interaction.[8] In the example discussed here, the vertical dotted lines in Graph 1 below mark the exact moment of the screenshot in Figures 1 to 4 . The curve starting with Fig. 4 shows three peaks (corresponding to the upward movements of Anna's head) coinciding with the three main accents of the chorally produced stretch of speech. Anna's head movements thus in essence 'conduct' the performance. We return to this vocal and postural synchronization in §6.3.

Graph 1. VEO-V „Orangensaft"

---
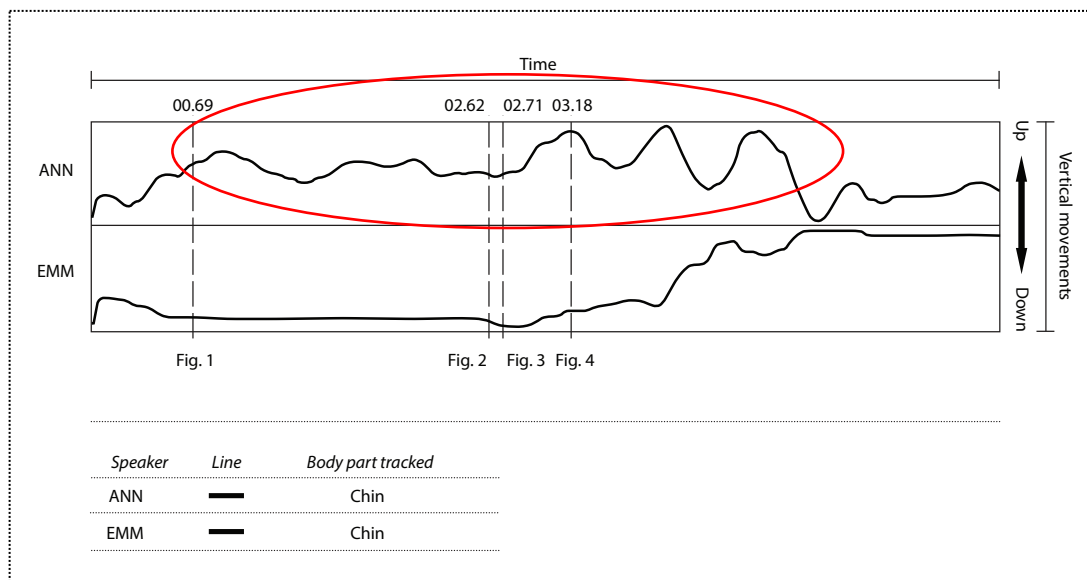
[8] VEO has been developed to adapt the advantages provided by the Motion Tracking Tool BLENDER to the needs of a conversation analytic analysis of interactional data. The core of BLENDER's tracking abilities is the *libmv library*, which includes, among other implementations, a differential method for estimating optical flow. An advantage lies in the fact that it is more precise than other point-wise methods since it is less sensitive to image noise and can hence resolve any inherent ambiguity. VEO uses BLENDER technology and adds the possibility to visualize a selected motion trajectory with the speech signal as displayed in PRAAT.

Extract (3) is in many ways the prototype of a choral performance. It is a bilateral accomplishment, with both participants equally active in bringing off the simultaneous speech. They are clearly working together to say the same words with the same prosody (intonation, rhythm and dynamics) in chorus. The words may be ones that derive from a shared socialization, including advertisement discourse, or ones that the participants have heard or said themselves on a prior occasion.[9] They serve to evoke the memory of a shared experience. Following their performance, the participants mutually ratify its success (lines 14 and 15) before turning to a new, hic-et-nunc topic.

4.2 Choral performance in French interaction

Turning now to French interaction, we find a similar prototype of choral performance as in German (Extract 3 above).

Adèle (on the right) and Inès (on the left) are sisters. This extract is not part of a 'joint remembering' of a past event but rather involves thinking about a similar event in the near future. Not long before the recording, Adèle and Inès had organized a big family reunion together. Here, they agree on doing a sort of 'debriefing' in order to think about what went well and what could be improved for a subsequent family meeting.

(4) "De jouer" 'to play'

```
408 Inès:   aprÈs ben on a attendu [les invit_-)]
            afterwards well we waited for the gues-

409 Adèle:                         [°h    <<f>ah] NON non non;>=
                                         oh no no no

410 Inès:   =quoi,
            what

411         (0.2)

412 Adèle:  je te coupe–
            I'm interrupting

413         (0.8)

414 Adèle:  tu SAIS,
            you know

415         (0.3)

416  Adèle: si vraiMENT on pouvait (en)fin;
            if we really could anyway
```

```
417          (0.3)

418 Adèle:   amélioRER,
             improve

419          (0.2)

420 Inès:    mhm,

421 Adèle:   puisqu'on est dans le déBRIEfing;
             since we're into debriefing

422          °hh on aurait PU quand même demander=
             we actually could have asked

423          =à TOUS les musiciens de notre faMILLE
             all the musicians in our family

424          (0.9)

425 Inès:    [de jouER;]
             to play

426 Adèle:   [de jouER;]
             to play

427          (0.9)

428 Inès:    ((click))

429          (0.5)

430 Inès:    <<p>ouais ça aurait été BIEN;>]
             yeah that would have been good
```

Adèle suggests that they could have talented family members bring their instruments to perform the music themselves instead of paying professionals to do it: *on aurait PU quand même demander à TOUS les musiciens de notre faMILLE* 'we actually could have asked all the musicians in our family' (lines 422 and 423), a syntactic trajectory that in French clearly projects completion, and then makes a longish pause of 0.9s (line 424). This moment of silence is followed by the choral production of the infinitival completion, *de jouER* 'to play' (lines 425 and 426). Inès then confirms their choral production with a click.

If we look at the bodily movements in the preparation phase of the choral production, we find that once again hand gestures play a decisive role, as in the previous example. The main speaker, Adèle, on the right, attracts her sister's attention with a fast forward-directed hand movement (Fig. 5), which subsequently turns into a held hand gesture, and with prolonged mutual gaze (Fig. 6). It is Inès, on the left, who first prepares to start speaking: her lips have already assumed a position for the articulation of the vowel /ə/. The choral performance starts just before Figure 8: *de …*, the speakers are perfectly synchronized when

uttering … *jouER* (Fig. 8). The line graph under Fig. 8 shows that the exact same wording is synchronized with very similar head movements.
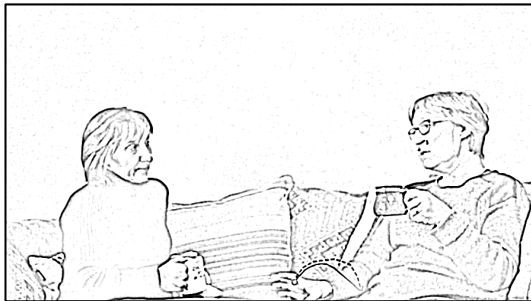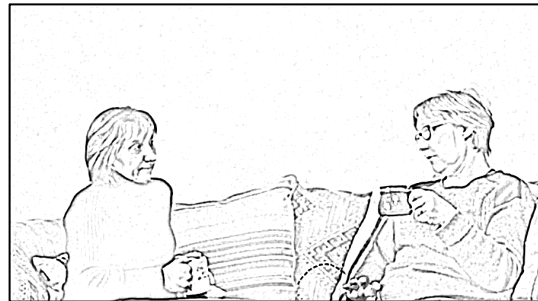
Fig. 5 Alert

Fig. 6 Hold

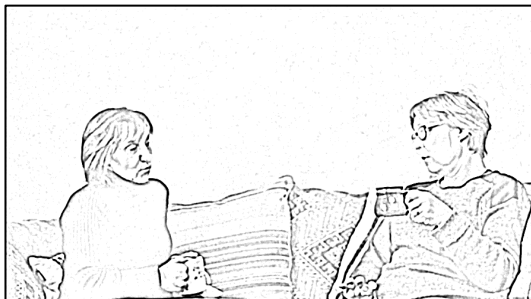

```
423 ADE:  =à TOUS# les musiciens de notre fa#MILLE-
```
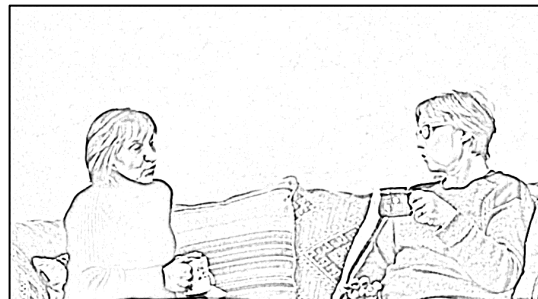
Fig. 7 Fine tune

Fig. 8 Start


```
425 INE:   [0.9#]                  425 INE:   [de #jouER;]
426 ADE:   [0.9#]                  426 ADE:   [de #jouER;]
```

In this case we tracked the head and hand movements of the two sisters in the vertical dimension. Graph 2 VEO-V "De jouer" shows that the sister on the left, Ines, is practically inert and in a prototypical bystander position up until Fig. 8, whereas the sister on the right, Adèle, as the main speaker in this sequence moves her head and her hands a lot.

Graph 2. VEO-V "De jouer"
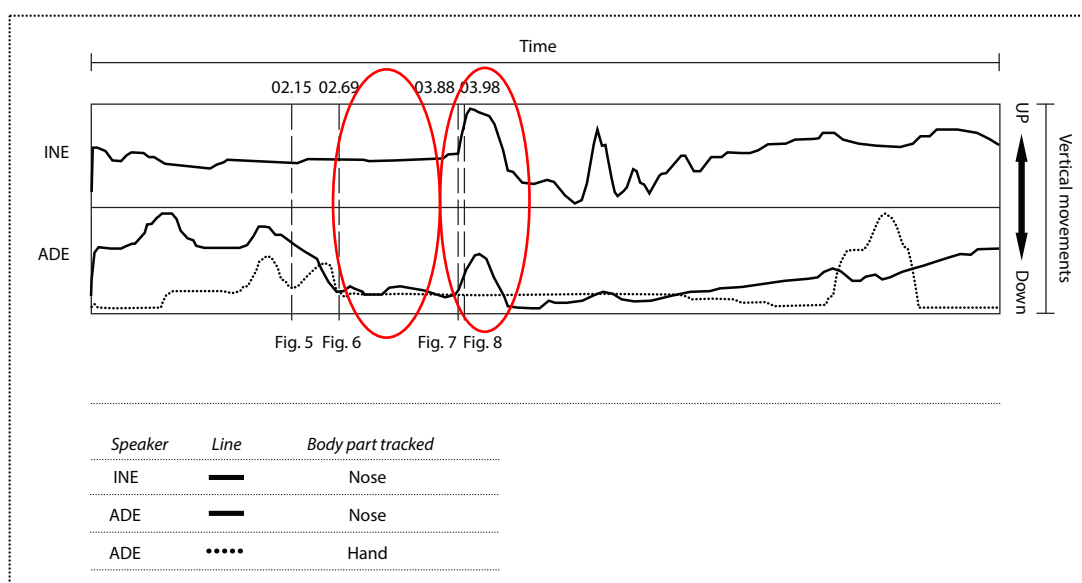
Interestingly, some time before the simultaneous start of the stretch of choral speech, the two participants go into a state of inertia; the lines between Figure 6 and Figure 7 in Graph 2 are almost parallel (see the left oval). The stretch of choral speech shows a strikingly similar gestalt in the up- and down-movements of the two sisters' heads (the right oval in Graph 2).

In sum, in both German and French interaction, choral performances are brought off through the projectability of what is about to be said (with varying strength of syntactic projection) and a fine temporal coordination facilitated by three preparatory phases (alert, hold, fine tune) prior to the choral production itself. Both participants work to achieve the performance and they display their appreciation of it at its closure.[10]

5. The prototype of chiming in

We turn now to a second prevalent type in our collection of simultaneous speech, one we have chosen to call *chiming in*. In contrast to choral performance, which comes about bilaterally through the concerted efforts of both participants, chiming in is a unilateral process. It is brought off primarily by the work of the second speaker, who closely monitors

---

[10] Note that the kind of appreciation provided is sensitive to what is currently being done by the interactants: in the example *Orangensaft*, an out-of-the-blue suggestion to have a cold beverage right now leads to a side sequence about snacking, whereas in the example *De jouer*, the appreciation takes the form of accepting the latest proposal for a future family reunion.

the current speaker to identify an appropriate moment to join in. Typically, what is co-produced is less than a TCU and occasionally less than a word. The incoming speaker's speech is softer in volume than that of the current speaker. In the aftermath, the current speaker typically continues the turn without acknowledging the co-participant's incoming.

5.1 Chiming in in German interaction

Here is a case from German interaction which is in many ways a prototypical instance of chiming in. The two participants, Anna and Emma, are the same as in Extract (3) above. Here they are reminiscing about a trip they took together with their friends, Martina and Manfred.

(5) „-lich nicht" '-ly not'

```
01 Emma:    ah weißt du noch dass da doch irgendwie WIEder:
            oh do you remember how there were again somehow

02          welCHE:
            those

03          ÄH:
            uh

04          verwirrungen WAren mit der pension
            mix-ups with the guest house

05          und dann eigentlich martina und manfred in EInem
            zimmer hätten [schlafen sollen]
            and then actually Martina and Manfred were supposed to sleep in one room

06 Anna:                  [°h STIMMT     ] (ja)
                           right               yeah

07 Emma:    und dann ist das so
            and then it is like

08          OH:,
            oh

09          HM,
            hm

10→         wOllten wir EIgent[lich NICHT? ]
            wanted we actually not
            (that's) not what we wanted actually

11→Anna:                      [lich NICHT- ]
                               –ly not

12 Emma:    aber dann haben die schon noch zwei [EXtra: zimmer  ]
            but then they still had two extra rooms

13 Anna:                                         [die hatten dann]
                                                  they got
            zwei geTRENNte zimmer ja
```

15

two separate rooms then yeah

Starting in line 01, Emma asks her friend Anna whether she remembers a specific episode (*weißt du noch* 'do you remember') and starts to narratively reconstruct this episode. Anna aligns in line 06 by uttering *STIMMT* 'right', thus contributing to making relevant the actuation of their shared experience (cf. Betz 2015). In line 07, Emma launches an animation of what Martina and Manfred reportedly said upon hearing that they would be sharing a room: *OH:, HM, wOllten wir EIgentlich NICHT*, colloquially 'that's not what we wanted actually' or more literally '(that) wanted we actually not', with the expression *eigentlich nicht* 'actually not' projecting a main accent on *NICHT* (lines 08-10). It is this expression that Anna aims to co-produce; she comes in on the final syllable of *eigentlich* and then produces *NICHT* in chorus with Emma but at a reduced volume (line 11). Emma does not acknowledge the co-production but instead continues the story (line 12).

There is a preparatory phase preceding the simultaneous speech in (5) in which the incoming speaker, Anna, on the left, begins to closely monitor the current speaker's talk, moving her head slightly closer to her friend (see Figures 9 and 10, and Graph VEO-H). Anna continues monitoring and moving forward – now not only her head, but also her torso (cf. Fig. 11), although Emma is looking away. Fig. 12 shows Anna launching her co-production just as Emma turns her head back to gaze at her. The co-production is thus brought off primarily by the incoming speaker, who waits until mutual gaze with the current speaker is re-established before coming in.

Fig. 9 Deep listening

06 ANN: [°h STIMMT] (ja)#

Fig. 10 Coming closer

07 EMM: dann ist das #so

Graph 3: **VEO-H**

Fig. 11 Still monitoring

08 EMM: #OH:

Fig. 12 Mutual gaze, chiming in

10 EMM: eigent**[#lich NICHT?]**
11 ANN: **[#lich NICHT-]**

Speaker	Line	Body part tracked
ANN	——	Nose
EMM	——	Nose
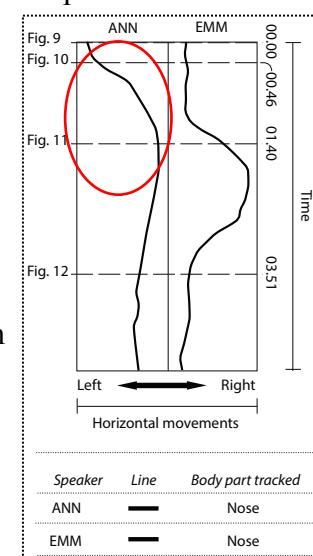
Simultaneous speech produced as chiming in is thus a unilateral accomplishment, resulting primarily from work by the incoming speaker, who closely monitors ongoing talk for an appropriate moment to join in. It is typically of short duration, limited to less than a TCU, and need not respect word boundaries. It is facilitated by strongly projective wording and prosody. There is no verbal or bodily acknowledgement or recognition that a co-production has occurred, nor any celebration of it in the aftermath: current speakers simply continue their ongoing project.

5.2   Chiming in in French interaction

Let us now consider an instance of chiming in in French interaction. As we will see, the prototype closely matches the German example above (Extract 5).

Pierre and Clara are married and have two daughters. In this extract, they are talking about their daughter Rachel's former boyfriends. They would have loved to see Rachel marry Daniel, which, however, never happened.

(6) "-lectuel", '-lectual'

```
01 Pierre:  (...)

02 Clara:    <<cresc> mai::s LUI euh,>=
                      but he uh

03           =oh_euh_<<cresc> il [était !PAR!]fait->=
              oh uh he was perfect

04 Pierre:                        [mais !LUI!,]
                                   but he

05 Clara:    =[il aurait été PARfai:t_il- ]=
              he would have been perfect_he

06 Pierre:    [(il serait/ il c'é- rait:)-]
               he would be he w-

07            [ben oui]_il était c'était le grAnd intélle[ctu][EL,][c' ]=
              well yeah      he was   it was the grand intellectual

08 Clara:    =[était  ]                                  [lec][tu][EL,]
               was                                        lectual

09           =[<<creaky> euh>;          ]
                         uh

10 Pierre:   =[était le gArs qui a f]ait (.)
              it was somebody who had done

              qui a été [à (.) harVARD, et qui a-]
              who had been to (.) Harvard and who had
```

```
11 Clara:                 [ah mais ça nous plaît çA] nous plaiSAIT,=
                           oh but we like that              we liked that

12            =ça nous plaisait BIEN-
               we really liked that

13            °h  (.) on a on avait pensÉ que peut-être que que
             ça [marcheRAIT;]
                 (.)  we thought we had thought that maybe that that it would work

14 Pierre:    [mmh mh m   ]_OUAIS;
                            yeah
```

Both Clara and Pierre work to list assessments of how their would-be son-in-law could best be described. First, Clara utters *il était PARfait* 'he was perfect' (line 03), then repairs her utterance to *il aurait été PARfa:it* 'he would have been perfect' (line 05), which is then echoed by Pierre (line 06). When Pierre now starts a syntactic trajectory *il était*, repairing to *c'était*, Clara chimes in on the last two syllables of the larger unit *le grAnd intéllectuEL* 'the grand intellectual' (lines 07 and 08). Notwithstanding Clara's intervention, Pierre continues to add elements to the list by stating that Daniel was someone who had been to Harvard and so on.

As in the German excerpts, chiming in comes about when the second speaker, here Clara, monitors the ongoing talk of the current speaker, here her husband Pierre. In Fig. 13, we can see that eye gaze is not necessary for close monitoring: an increase in attention can be detected due to a change in facial expression and body tonus. If we look at Figures 13, 14, and 15 sequentially, we can follow Clara's eye gaze gradually looking over to her husband. During this preparatory phase, Clara shifts slightly closer to Pierre (Fig. 14 and Graph 4). Fig. 15 shows Clara fixing her gaze on Pierre and his ongoing speech production. At the precise moment when Pierre glances at her and mutual gaze is established, she joins in on his syntactic trajectory, producing only the last syllables *-lectuel* of *intéllectuEL* with a slight delay.

Fig. 13 Monitoring ('deep listening')**Fig. 14** Coming closer   **Graph 4. VEO-H**


```
07 PIE: [ben oui]il# était   #c'était le
08 CLA: [était  ]
```

Fig. 15 Still monitoring (visually)**Fig. 16** Mutual gaze, chiming in

```
07 PIE:        grAnd inté#lle#[ctu][EL,][c' ]=
08 CLA:                       #[lec][tu ][EL,]
```

In sum, chiming in comes off in both German and French interaction in a similar fashion. It is accomplished primarily through the work of the recipient, who intensely monitors the ongoing talk of the current speaker, often moving slightly closer. When a syntactic project is launched that permits projection of how it will continue, the recipient comes in, at a reduced volume, on a final word/expression or part thereof once mutual gaze is established. Typically, the current speaker does not acknowledge the simultaneous speech but continues to pursue the ongoing turn or project.

6.   Semiotic resources: How are choral performances and chiming in accomplished?

Both formats for simultaneous speech require a high degree of coordination between participants in order to succeed. Neither comes out of the blue; both are well prepared. In each type, the coordination is facilitated by multiple semiotic resources, including verbal, prosodic and bodily means. Yet what we wish to show in this section is that the two formats are brought about by utilizing these semiotic resources in different ways. Ultimately, the synchronization is accomplished through different *practices* with the same or similar semiotic resources.

6.1 Verbal projection

In both choral performance and chiming in, what the current speaker is in the process of saying allows the second speaker to anticipate what will come next. However, the scope of this

projection is wider in the case of choral performance. For instance, in Extract (3) "Frisch gepressten Orangensaft", Anna first mentions the orange juice in the Spanish café and then builds a turn referring to it *da hätt ich jetzt Lust drauf* 'I could go for that now', literally 'there could I now go for that'. This turn evokes an imaginary world in which she could have a glass of Spanish orange juice. It is followed by a lengthened inbreath, a signal projecting imminent talk (Hoey 2014), and a reference to the object of desire (the inbreath is hearable as a (longing) sigh here). Emma can thus anticipate what this will be and, in unison with Anna, she produces a cliché-like referential form, one they are likely to have seen, heard, and used on other occasions: *frisch gepressten Orangensaft* 'freshly squeezed orange juice'.

In the case of chiming in, the second speaker must also be able to anticipate what the ongoing speaker is about to say, but here the scope of projection is much narrower. In (5) "-lich nicht", for instance, Anna can anticipate from the way Emma has set up the story that Martina and Manfred will be reported as saying they did not want to share a room. In other words, the negative particle *nicht* 'not' is strongly projected. However, Anna cannot know until the very last moment that Emma will insert the optional adverb *eigentlich* 'actually' before this particle. She thus joins in late, co-producing only the last syllable of *eigentlich* before reaching the projected *nicht*.

6.2 Prosodic projection

In both choral performance and chiming-in formats, second speakers orient to main accents that are projected by the ongoing talk of the current speaker. In addition, however, with choral performances, especially German ones, a strong isochronous rhythm is noticeable in the current speaker's talk prior to the onset of simultaneous speech, the isochrony being created by accents placed at regular intervals in time. It is as if the current speaker is setting up a metronome for the second speaker to orient to when joining in. This is, for instance, especially noticeable in (3) "Frisch gepressten Orangensaft". Here Anna, who first introduces the topic of orange juice with *och der Orangensaft da hätt ich jetzt Lust darauf* 'oh the orange juice I could go for that now', places primary accents on *OCH* and *oRANgensaft*, a secondary accent on *hÄtte,* and another primary accent on *LUST,* which are timed to come roughly 0.4 secs apart: see Figure 17, where the intervals between the accented syllables are shown in Praat

pictures and colored blue.[11] The slashes in the text beneath each Praat picture mark off the stretch of speech being produced in each interval.

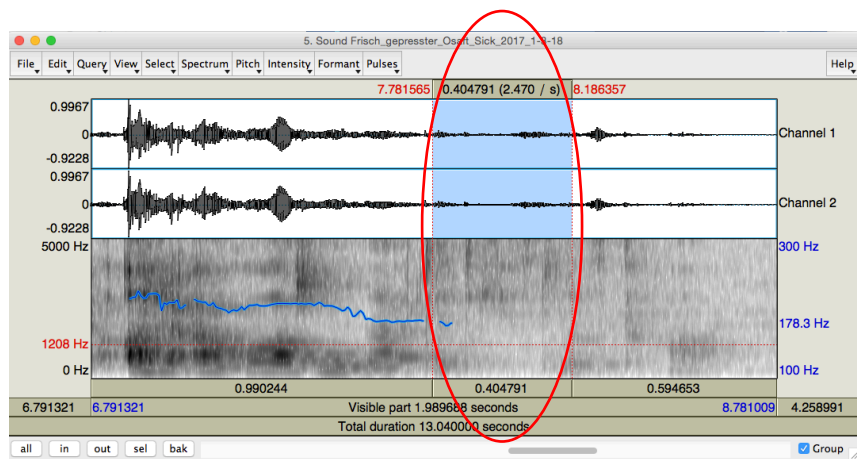**Figure 17.** Isochronous intervals established by strong accents in Extract (3)[12]
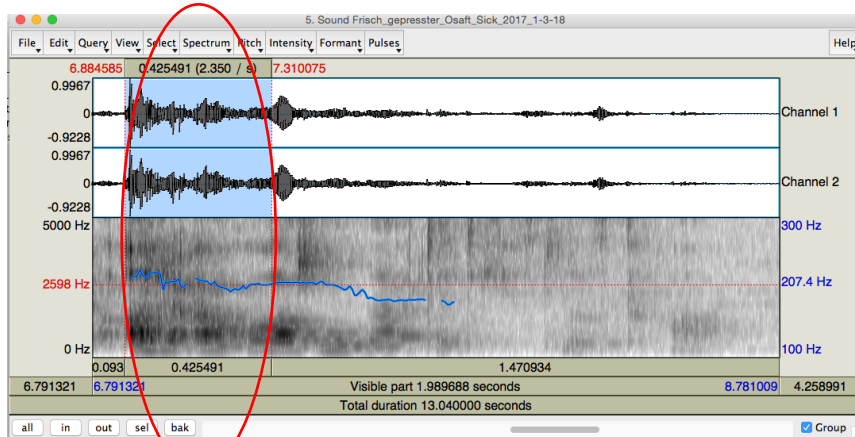
/OCH der o/RAN-
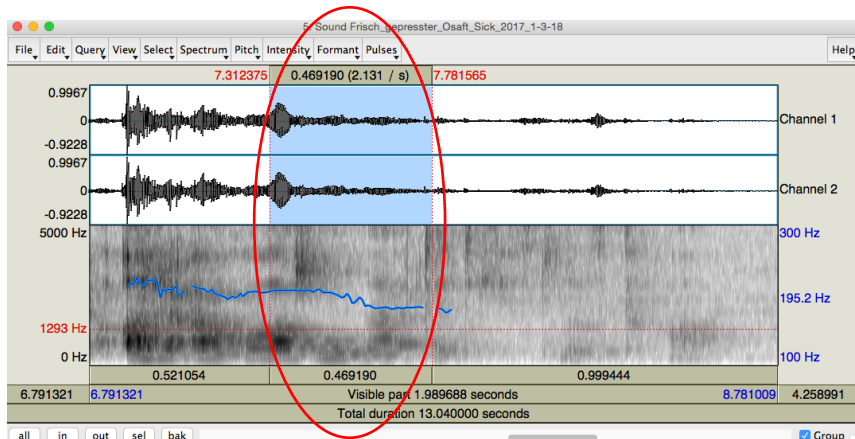
/RANgensaft=da /hÄtt

/hÄtt ich jetzt /LUST

The regular intervals created by the timing of these accented syllables create a metric for the timing of subsequent talk, which – following an inbreath by Anna -- is a stretch of speech that Emma joins in on to produce simultaneously with Anna: (*oh so_n/ah ja (n)) frisch gepressten Orangensaft* '(oh like a/oh yeah (a)) freshly squeezed orange juice'. This stretch of speech is also strongly rhythmic, with accents coming on *frIsch, gePRESSTen, and oRANgensaft* at regular intervals in time. Interestingly, however, Anna adopts a slightly slower tempo (beats every 0.6 sec.), which forces Emma to re-adjust slightly in order to make the synchronization perfect. Yet it is the rhythmic framework established in Anna's prior talk that facilitates the temporal coordination of this choral performance.

In French choral performances, the rhythmic projection is syllable-based rather than accent-based.[13] Nevertheless, we can observe regular prosodic phrases being created in the talk immediately preceding simultaneous speech here too, the regularity being created by isometry, whereby intervals contain roughly the same number of syllables of equal duration (Auer et al. 1999). These isometric intervals serve to project when the syllables of the next prosodic phase are due. Extract (4) "De jouer" is a particularly interesting case of such projection. Adèle produces two prosodic phrases with roughly the same number of syllables and of equal duration at the beginning of her turn: *on aurait PU quand même demander à* (10 syllables) 'we actually could have asked' and *TOUS les musiciens de notre faMILLE* (9 syllables) 'all the musicians in our family'.[14] The duration of these phrases (1.4 sec. each) projects that the next prosodic phrase should be equally long; indeed, this is what allows Adèle and Inès to know that they should wait 0.9 sec. before co-producing the next three syllables *de jouER* 'to play'. The pause

---

[13] For a discussion of accent-based vs. syllable-based timing in languages see Auer et al. (1999).

[14] The utterance so far now syntactically projects more to come.

and these three syllables together create a prosodic phrase of the same duration as the previous two (see Fig. 18).
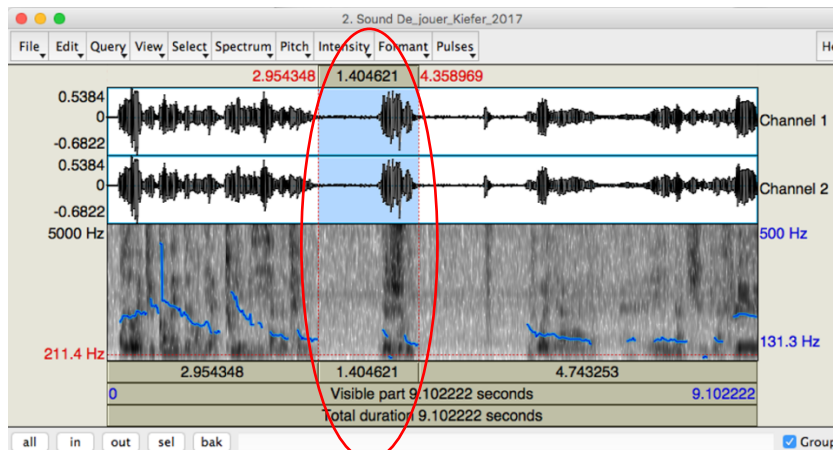
Figure 18. Isometric prosodic phrases in Extract (4) marking off regular units in time
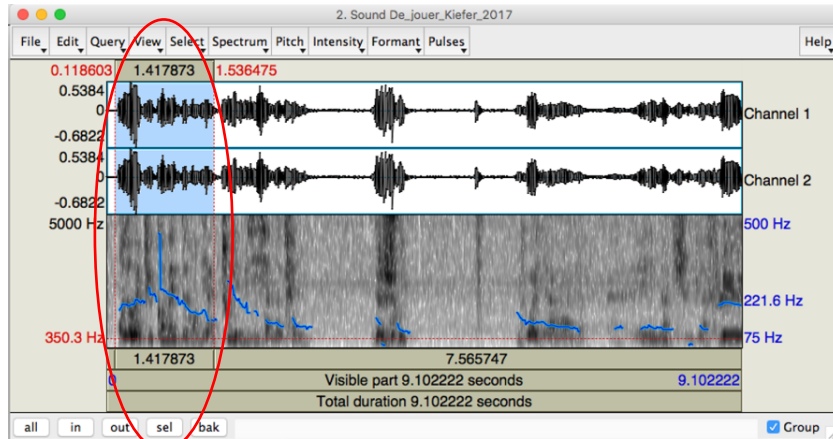
/on aurait PU quand même demander à/
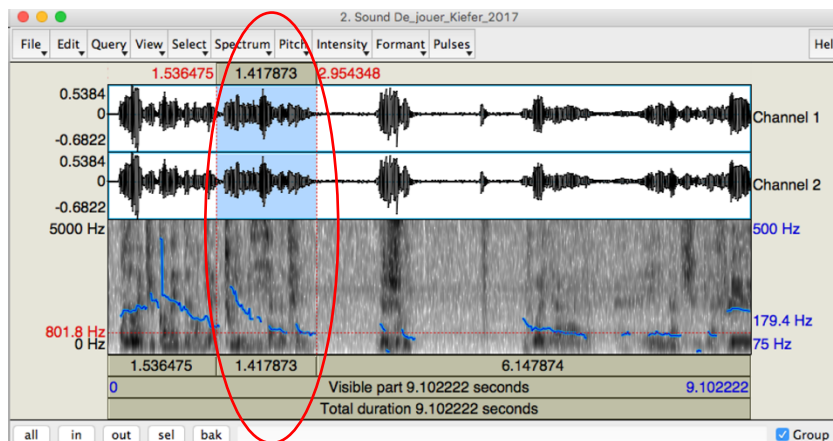
/ TOUS les musiciens de notre faMILLE/

/ (0.9) de jouER/

With chiming-in formats, by contrast, strong rhythmization is not as common in a current speaker's talk prior to the onset of simultaneous speech.[15] Instead, incoming speakers time their simultaneous speech to coincide with the main accent projected in a current speaker's ongoing talk.

Prosodic projection then can be observed in both formats. Whereas in the case of chiming in, second speakers orient to a projected main accent, in choral performance, projection is strongly rhythmic (either isochronous or isometric): the current speaker can be seen to set up a rhythmic frame, either with regularly occurring accents or with a similar number of regularly timed syllables, thereby facilitating the timing of the co-production. Second speakers use this frame to coordinate their incoming with ongoing talk. In this sense, the accomplishment of simultaneous speech is bilateral. Chiming in, on the other hand, represents a unilateral accomplishment: it is primarily incoming speakers who use prosodic projection in ongoing talk to anticipate an upcoming main accent and time their incoming to coincide with it.

6.3 Bodily practices in choral performance and chiming in

In this section, we examine the interplay of embodied means as used in the preparation, implementation, and aftermath of simultaneous speech. We aim to show that the two formats are brought about by utilizing the same semiotic resources in different ways immediately before (§6.3.1), during (§6.3.2) and immediately after simultaneous speech (§6.3.3).

6.3.1   Preparing turn sharing

In the preparatory phase of choral productions, an alert is first performed (often through an other-directed hand gesture: see Fig. 19a).

---

[15] At times, chiming in can display rhythmic gestalts, e.g., if the simultaneously uttered stretch of speech is part of a polyphonic animation (see Ex. 5). Note, however, that these exceptions hold only for the joint performance itself, not for the preparatory phase. There are two reasons for this. First, whereas choral performances are based on a joint preparation by both participants, chiming in is basically brought about by the incoming speaker alone. Second, whereas the jointly produced stretch of speech may contain several main accents in choral performances, it is generally only one main accent that is produced simultaneously in chiming in.

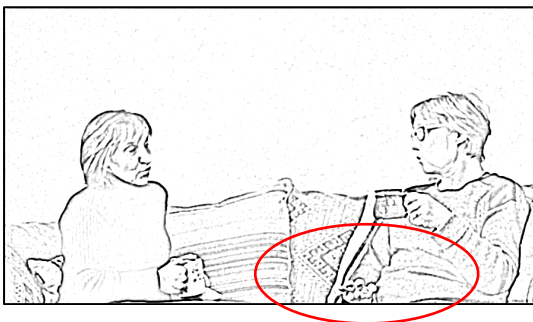**Fig. 19: Hand movement performed by current speaker on the right**

19a
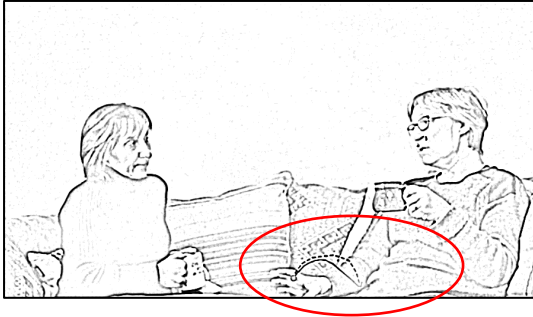
19b
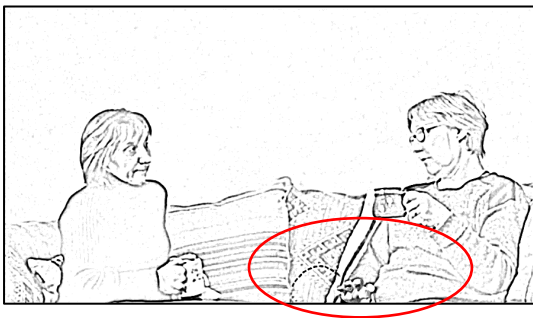
19c

Second, this gesture is withdrawn (Fig. 19b). Third, just before the choral onset, both participants fine tune their movements within a split second using a variety of different bodily resources while the hand is still held (Fig. 19c). Note that these three gesture phases correspond exactly to the vocal accents in *TOUS, faMILLE* and *jouER*; thus, a rhythmic gestalt is created with both vocal and bodily resources.[16]

---

[16] This observation, which suggests a striking similarity between German and French, is reminiscent of Szczepek & Persson's (2016) findings on glottalization at word boundaries in the two languages. The authors determine that the same phonetic contrast is used to indicate the nature of turn continuation in both languages and conclude that „sound patterns shape interaction" (i.e., are practices for organizing and managing talk, S.P. & E.C.-K.)

In chiming in, the three phases comprise first a high level of attention, which is displayed through monitoring (Fig. 20).

Fig. 20: Incoming speaker monitoring current speaker

Second, the participant who is about to join in moves slightly forward or turns towards the current speaker, coming slightly closer to their interlocutor (Fig. 21):

Fig. 21: Incoming speaker coming closer to current speaker
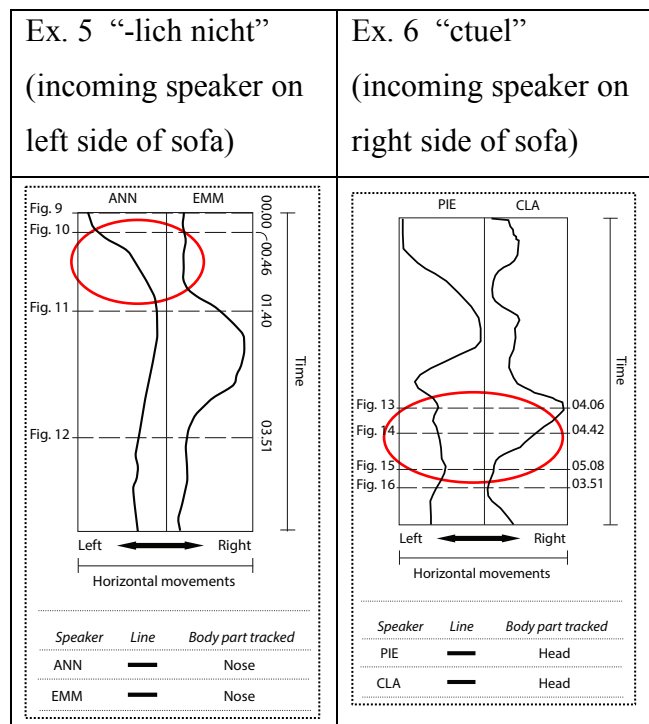
„and are not solely determined by language-specific phonologies" (p. 128). More work is clearly needed to explore the import of such cross-linguistic commonalities.

The third phase is best described as "still monitoring", up until the current speaker redirects their gaze towards the incoming speaker, whereupon simultaneous speech begins.

In a nutshell, the preparation phase of chiming in builds on monitoring, mostly via *gaze*, and on a change of posture by one participant ('coming closer'), whereas choral performances involve a high degree of complex bodily synchronization by both participants.
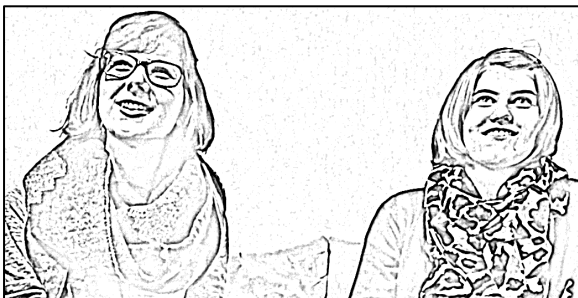
6.3.2 During simultaneous speech

A high level of bodily synchronization is also observable during the choral performance itself. To illustrate this observation, the next two screenshots are taken from Ex. 3 "Frisch gepressten Orangensaft". Note that the bodily and facial expressions are aligned with the main accents of the verbal utterance. Both facial and bodily expressions are also highly synchronized between the two interactants, who at the exact same moment are raising their heads while appreciatively closing their eyes and lips (Fig. 22) and subsequently opening their eyes and lips in a smiley, longing expression while looking to the ceiling (Fig. 23):

Fig. 22 Synchronized bodily expression on /-prEss-/

Fig. 23 Synchronized bodily expression on /-RAN-/

With chiming in, this kind of facial and postural synchronization is not observed.

6.3.3   In the aftermath of simultaneous speech

Let us now examine what happens immediately after a choral performance. The following screenshots, taken from "Frisch gepressten Orangensaft" (Ex. 3: Fig. 25) and "De jouer" (Ex. 4: Fig. 26), show how the two sets of friends perform a gestalt closure through jointly smiling or grinning.

Fig. 25 The aftermath of "Orangensaft"   **Fig. 26.** The aftermath of "De jouer"

The gestalt closure in other choral performances is achieved through jointly produced head nods or a click that multimodally closes the gestalt.[17] Moreover, the co-produced stretch of talk is typically followed by both speakers changing their posture (thereby publicly displaying the closure of a shared activity). In chiming in there is no acknowledgment of the simultaneous speech by the current main speaker and therefore no collaboratively produced gestalt closure afterwards.

Our findings can be systematized as shown in Tables 1 and 2.

---

[17] For gestalt closings in collaborative utterances cf. Oloff 2013.

Table 1. **Bodily practices in** CHORAL PERFORMANCES

PHASES	ALERT	HOLD	FINE TUNE	*CHORAL SPREECH*	CLOSING GESTALT
Orangensaft Ex.3, German	Break in flow of bodily movement (raising head and turning away)	Inbreath plus hold	Fine tuning of partners' rhythm/tempo	*Almost identical facial expressions and body posture on main accents*	Click, closed eyes and head nods; smiling
De jouer Ex.4, French	Break in flow of bodily movement (hand gesturing)	Holding gesture	Fine tuning of joint speech onset	*Almost identical head movement and change in body posture on main accent*	Click, closed eyes and head nods; grinning
Speaker agency	*Current speaker*	*Current speaker*	*Both*	*Both*	*Both*

Table 2. **Bodily practices in** CHIMING IN

PHASES	MONI-TORING	COMING CLOSER	STILL MONI-TORING	*CHIMING IN*	BACK TO BUSINESS
-lich nicht Ex. 5, German	Deep listening to current speaker	Turning head, coming forward and closer	Waiting for mutual gaze	*Incoming speaker holds gaze*	Back to monitoring position
-ctuel Ex. 6, French	Deep listening to current speaker	Coming closer, turning head	Waiting for mutual gaze	*Quick mutual eye contact*	Turning back to start position
Speaker agency	*Incoming speaker*	*Incoming speaker*	*Incoming speaker*	*Both*	*Incoming speaker*

The most striking difference to emerge from the tables above is that while choral performances (Table 1) are prepared through hand movements and holding gestures by the current speaker, chiming in (Table 2) is prepared for through eye gaze monitoring or deep listening by the incoming speaker plus a slight posture adjustment that allows for better monitoring. This difference can be accounted for by assuming that it is the current speaker who takes the initiative in choral performances, with the two speakers collaborating during simultaneous speech, while in chiming in it is the second speaker who takes the initiative, one that is not explicitly taken up by the first speaker.

What we hope to have shown in this section is that the two formats, choral performance vs. chiming in, are brought about by utilizing the same semiotic resources in different ways for preparing, performing and bringing to an end the sharing of turns. The practices associated with these formats are clearly differentiated not only with respect to who takes the initiative (current vs. incoming speaker), but also with respect to how participants make use of bodily resources in executing the simultaneous speech itself.

7. Functional aspects: What do choral performances and chiming in allow participants to achieve?

We have argued so far that the two formats for simultaneous speech are brought off either in a bilateral or a unilateral process. In this section, we wish to explore what participants achieve by designing their speech to be simultaneous. As might be expected, the two different formats are quite distinct in the interactional effects they have.

7.1 Choral performance: Display of a shared affective stance

In our data, participants talk collaboratively about events they have experienced in the past. Strikingly, though, in all the choral performance instances, the stretch of talk produced collaboratively is not about reconstructing something in the past. Rather, the two participants are publicly displaying through multiple semiotic channels an affective stance (cf. Niemelä 2010) towards something in an imagined, often better world than the one they have experienced. We see this happening, for example, in "Frisch gepressten Orangensaft" (Ex. 3); the close friends are talking about their breakfast experience during a holiday in Spain and the

freshly squeezed orange juice they had there. Reminiscing about this experience suddenly inspires one of the co-tellers to imagine how wonderful it would be to enjoy a similar cold beverage right now. Similarly, the sisters remembering a past family reunion in the extract "De jouer" (Ex. 4) switch from reconstructing the evening they experienced together to considering what they could do to make it even better the next time. Note, however, that the sisters do not merely rationally list better options; rather, they imbue these options with feeling, imagining how joining their musical skills would lead to an even greater experience of family togetherness.

In a nutshell, we find roughly the same interactional function for choral productions, viz. displaying an affective stance towards something in an imagined world. The co-interlocutors display not only a shared understanding of this alternative world, but also an affective, often positive stance towards something in that world. The two participants are full agents in a tightly coordinated performance that evokes sharedness. Their display of shared appreciation is meant not only for the third person in the room (and/or the camera), but also for each other.

7.2 Chiming in: Making an epistemic claim to know just as much

In all cases of simultaneous speech in our collections, the two participants have equal epistemic rights to know about the events they are talking about.[18] Yet invariably one of the two takes the initiative in recalling or remembering particular details. We see this happening, for instance, in (6) "-ctuel" when Pierre and his wife Clara are describing the high esteem they had for their daughter's would-be suitor Daniel. Initially, there is some competition between the two participants as to who will be speaker (lines 03-06) but Pierre prevails by agreeing with Clara's evaluation that Daniel would have been 'perfect'. He then begins an enumeration of his qualities: *c'était le grand intéllectuel* 'it was the great intellectual' using list intonation. When Clara now chimes in with Pierre to co-produce the last two syllables of *intéllectuel*, she is thus also making an epistemic claim (Heritage 2012) to know (and agree) that one of the things that made Daniel perfect was that he was an intellectual. The fact that she gazes at Pierre ever so briefly when chiming in suggests that her claim is directed as much to him as it is to the third person in the room.

---

[18] In this, the practice of chiming in comes close to 'collaborative completions' in that both interactants ephemerally "share the role of speaker for the time of a conversational topic on which they have mutual knowledge" (Szczepek Reed 2000: 2; cf. also Falk 1980, Günthner 2013, among others).

The same pattern is found in other cases of chiming in: in (5) "-lich nicht", for instance, Anna displays that she too knows what the couple said at the hotel reception. Yet because of the way chiming in comes off, the incoming speaker is not challenging the epistemic authority of the current speaker. Instead, s/he is merely adding – *sotto voce* – a display of 'I know that too', thus supporting an earlier epistemic claim (cf. Anna's *stimmt* 'right' in line 06). This may account for why the current speaker does not acknowledge or otherwise react to chimed-in simultaneous speech: the initiative of the turn remains with them. Only the current speaker is a full agent, with the incoming speaker exercising a supportive role at most.[19]

8.   Discussion and conclusion

One of our initial interests in examining simultaneous speech in a language other than English was to explore how the affordances of a specific language might affect turn-sharing (see also Hayashi 2013:188). In this article we have explored simultaneous speech in a data set of French and German. Although our collection contains many more examples, we have only been able to deal with a small sample of cases from each language. It goes without saying that our conclusions are preliminary to the extent that they are based on this sample.

8.1 Language-specific inflections of simultaneous speech

Comparing co-productions in a Germanic language (German) and a Romance language (French) seemed particularly promising since their rhythmic phonologies are diametrically opposed: the 'metronome' of spoken interaction in German is accent-based, while for French it is syllable-based (for more on this distinction, see Auer et al. 1999). Consequently, we initially hypothesized that choral performances would be found only in German, which allows for far-ranging rhythmic projection, and chiming in only in French, where rhythmic projection was believed to be more local. Careful examination of the data in our collection of 32 turn sharings, however, proved us wrong.

In fact, both French and German allow for the possibility of choral performance, just as both allow for chiming in. Of course, their specific linguistic resources – verbal and

---

[19] Although in the examples shown here the incoming speakers enter in the middle of words, this is not invariably the case. In other examples in our collection, the incoming speaker chimes in with whole words.

prosodic, in particular – are distinct. However, as we have tried to show, the way these different resources are deployed as practices for implementing choral performance and chiming in are quite similar across the two languages (see also Auer & Pfänder 2008 and Szczepek Reed & Persson 2016 for similar observations in comparing French and German conversational data). Speakers of both languages make use of far-reaching, 'global' verbal and prosodic/rhythmic projection to facilitate choral performance, whereas they deploy 'local' verbal and prosodic projections for chiming in.

Moreover, as our micro-analysis made us increasingly aware, simultaneous speech also involves bodily resources, to which speakers of both languages have the same access. In fact, it emerged from our data that independent of which specific language they were speaking, our participants were deploying their body in similar ways to implement the two distinct formats for simultaneous speech.

In both languages, a choral performance in our data comes off bilaterally through the concerted efforts of the two participants, while chiming in in our data is accomplished unilaterally through the careful monitoring of the current speaker's talk by the incoming speaker. With these different practices, our participants are able to achieve something similar in the two languages: a bilateral display of affective congruence and reciprocal appreciation vs. a unilateral (secondary) epistemic claim to similar knowledge and/or agreement with the current speaker's evaluation.

8.2 Turn-sharing revisited: two sedimented interactional practices for simultaneous speech

Our empirical analysis of simultaneous speech has built on Lerner's (2002) contribution. We follow Lerner in that we focus on those cases in which the turn-sharing does not occur by accident nor by competitive motivation. Rather, in all our examples simultaneous speech is achieved interactionally. We also subscribe to Lerner's claim that interactionally achieved simultaneous speech can have several interactional functions. Going beyond Lerner, however, we have argued that there are two distinct groups of samples in our collection. Each group corresponds to a sedimented practice in interaction that is perceivable to the participants. Both practices, choral production and chiming in, have been shown to have a specific three-step preparatory phase, as well as a typical way of being performed and a specific type of continuation after the co-production. Not only do the two practices have their own formal design but they also have their own characteristic interactional function. While chiming in

allows the speaker who joins in to publicly display an epistemic claim of 'I know that too', choral productions allow for the display of a collaborative affective stance towards an imagined and positively evaluated alternative scenario.

Going beyond a mere analysis of verbal production, our study reveals that participants build on bodily resources for the two practices we have distinguished. In choral performances, a co-interlocutor's attention is triggered through a gestural alert, which is often held as participants fine-tune their activity in order to collaboratively launch the simultaneous performance. For chiming in, eye gaze plays a more important role than hand movements, as the second speaker closely monitors the current speaker, often leaning forward slightly. If the first speaker's gaze is averted, the incoming speaker waits for their eyes to meet again before joining in on the ongoing syntactic project. Other bodily resources – some, such as audible and/or visible inbreath, which have already been described in the turn-taking literature, as well as others, such as slight changes in posture, which have only rarely been taken up in interaction studies -- have also been shown to play a constitutive role in orchestrating shared turns. We hope thus to have righted the picture of turn sharing as a purely verbal phenomenon.

References

Alcón, Daniel & Muz, Daniel & Satti, Ignacio & Lahmann, Claas (submitted): Visualising Emergent Objects (VEO). A new methodological workflow for the multimodal analysis of conversational data.

Atkinson, J. Maxwell (1984). Public speaking and audience responses: Some techniques for inviting applause. In: *Structures of Social Action: Studies in Conversation Analysis*, ed. by J. Maxwell Atkinson and John Heritage. Cambridge, Cambridge University Press: 370-409.

Auer, Peter (1990). Rhythm in telephone closings. *Human Studies* 13, 361-392.

Auer, Peter & Frank Müller & Elizabeth Couper-Kuhlen (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. New York and Oxford, Oxford University Press.

Auer, Peter & Stefan Pfänder (2008). Les allers-et-retours sur l'axe syntagmatique en allemand et en français parlés. Une étude contrastive de rhétorique orale. *Cahiers de Praxématique* 48, 57-84.

Betz, Emma (2015). Indexing epistemic access through different confirmation formats: Uses of responsive *(das) stimmt* in German interaction. *Journal of Pragmatics*, 87, 251-266.

Betz, Emma & Golato, Andrea (2008). Remembering relevant information and withholding relevant next actions: The German token *achja*. *Research on Language and Social Interaction*, 41(1), 58-98.

Bolden, Galina, Hepburn, Alexa, & Potter, Jonathan (2019, in press). Subversive completions: Turn-taking resources for commandeering the recipient's action in progress. *Research on Language and Social Interaction*, 52(2).

Couper-Kuhlen, Elizabeth & Barth-Weingarten (2011). A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, Margret et al.: Gesprächsanalytisches Transkriptionssystem 2. *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion*. 12: 1-51. www.gespraechsforschung-ozs.de

Couper-Kuhlen, Elizabeth (2018). Finding a Place for Body Movement in Grammar. *Research on Language and Social Interaction*, 51:1, 22-25.

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

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

Cummins, Fred (2014). The remarkable unremarkableness of joint speech. In *Proceedings of the 10th International Seminar on Speech Production*: 73-77.

Cummins, Fred (2018). *The Ground From Which We Speak*. Fred Cummins (published online).

Drew, Paul (2009). "Quit talking while I'm interrupting:" (a comparison between) positions of overlap onset in conversation In Markku Haakana, Minna Laakso, & Jan Lindström (Eds.), *Talk in Interaction: Comparative Dimensions* (pp. 70-93). Helsinki: Finnish Literature Society.

Ehmer, Oliver (2011). *Imagination und Animation: die Herstellung mentaler Räume durch animierte Rede*. Berlin: Walter de Gruyter.

Falk, Jane (1980): The Duet as a Conversational Process. *University Microfilms International*. Ann Arbor, Michigan.

French, Peter & Local, John (1983) Turn-competitive incomings, *Journal of Pragmatics* 7: 17-38.

Goffman, Erving (1978). Response cries. *Language* 54: 787-815.

Golato, Andrea (2012). German *oh*: Marking an emotional change-of-state. *Research on Language and Social Interaction*, 45(3), 1-24.

Günthner, Susanne (2013). Ko-Konstruktionen im Gespräch: Zwischen Kollaboration und Konfrontation, *Grammatik in der Interaktion*, gidi-Arbeitspapiere 49.

Hayashi, Makoto (2005). Joint turn construction through language and the body: Notes on embodiment in coordinated participation in situated activities. *Semiotica* 156(1/4): 21-53.

Hayashi, Makoto (2013). Turn allocation and turn sharing. *The Handbook of Conversation Analysis*. Oxford: Wiley Blackwell, 167-190.

Hayashi, Makoto (2014). Activity, participation, and joint turn construction. In: Kabata, K., & Ono, T. (Eds.), *Usage-based approaches to Japanese grammar: Towards the understanding of human language* (Vol. 156). John Benjamins Publishing Company: 223-258.

Heritage, John (1984). A change-of-state token and aspects of Its sequential placement. In J. Maxwell Atkinson & John Heritage (Eds.), *Structures of Social Action* (pp. 299-345). Cambridge: Cambridge University Press.

Heritage, John (2012*). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction*, 45(1), 1–29.

Heritage, John, & Clayman, Steven E. (2010). *Talk in Action: Interactions, Identities and Institutions*. Oxford: Blackwell-Wiley.

Heritage, John & David Greatbatch (1986). Generating applause: A study of rhetoric and response at party political conferences. *American Journal of Sociology* 92(1): 110-157.

Hoey, Elliott M. (2014). Sighing in Interaction: Somatic, Semiotic, and Social. *Research on Language and Social Interaction* 47(2): 175-200.

Jefferson, Gail (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica* 9 (1), 47-96.

Lerner, Gene H. (1992). Assisted Storytelling: Developing Shared Knowledge as a Practical Matter. *Qualitative Sociology* 15(3): 247-271.

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

Lerner, Gene H. (1996). On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker. In: *Interaction and Grammar*, ed. by Elinor Ochs, E. A. Schegloff, Sandra A. Thompson. Cambridge, Cambridge University Press, 238-276.

Lerner, Gene H. (2002). Turn-sharing: the choral co-production of talk-in-interaction. In: *The Language of Turn and Sequence* ed. by Cecilia E. Ford, Barbara A. Fox and Sandra A. Thompson. Oxford, Oxford University Press: 225-256.

Niemelä, Maarit (2010). The reporting space in conversational storytelling: Orchestrating all semiotic channels for taking a stance. *Journal of Pragmatics* 42 (12), 3258-3270.

Ogden, Richard (2013). Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, 43(3), 299-320.

Oloff, Florence (2013). Embodied withdrawal after overlap resolution. *Journal of Pragmatics* 46: 139-156.

Pillet-Shore, Danielle (2012). Greeting: Displaying stance through prosodic recipient design. *Research on Language & Social Interaction*, *45*(4), 375-398.

Schegloff Emanuel, A., & Gail, J. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

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

Schwitalla, Johannes (1993). Über einige Weisen des gemeinsamen Sprechens. Ein Beitrag zur Theorie der Beteiligungsrollen im Gespräch. *Zeitschrift für Sprachwissenschaft*, *11*(1), 68-98.

Szczepek Reed, Beatrice (2000): Functional Aspects of Collaborative Productions in English Conversations. In: Interaction and Linguistic Structures (InLiSt) 21.

Szczepek Reed, Beatrice & Raymond, Geoffrey (eds.) (2013). *Units of Talk – Units of Action*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Szczepek Reed, Beatrice & Persson, Rasmus (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.

Tannen, Deborah (1989). *Talking voices. Repetition, dialogue, and imagery in conversational discourse.* Cambridge, Cambridge University Press.

Walker, Gareth (2016). Phonetic variation and interactional contingencies in simultaneous responses. *Discourse Processes* 53(4), 298-324.

Wright, Melissa (2011a). On clicks in English talk-in-interaction. *Journal of the International Phonetic Association* 41(02): 207-229.

Wright, Melissa (2011b). The phonetics–interaction interface in the initiation of closings in everyday English telephone calls. *Journal of Pragmatics* 43: 1080-1099.