



Non-lexical vocalisations + “so_was” as a multimodal package in establishing joint decisions in music rehearsals

Angeliki Balantani

Section d'allemand, University of Lausanne, Switzerland



ARTICLE INFO

Article history:

Keywords:

Multimodality
Conversation analysis
Deixis
Deontics
Multimodal package
Gaze

ABSTRACT

This study examines non-lexical vocalisations accompanied by the German deictic “so” or “so_was” as a multimodal package in music rehearsals. Focusing on the sequences prior to the performance of a piece, where this practice often occurs, I demonstrate that the multimodal package, vocalisations + “so” or “so_was”, is utilised by participants in the service of establishing a joint decision on how to perform the next part of a musical piece. The analysis shows that participants demonstrate the rhythm in which the next piece should be played and, by seeking their interlocutors’ approval in the next turn, orient to their shared deontic authority with regard to decisions about their joint performance.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Joint actions are part of everyday social life. Many activities involve the coordination of multiple participants for the successful accomplishment of a task. Moving a big cupboard, scoring a goal in a football match or even rock-climbing requires careful planning and coordination of the participants involved in the activity. In each of these activities, participants need to cooperate and often coordinate themselves on how to proceed in order to achieve their goals. This holds also for participants in music rehearsals. Before proceeding with the main task of playing their instruments, participants of a rehearsal session coordinate their actions and agree on how they are going to proceed, what rhythm they are going to follow or how quick the tempo is going to be. This coordination in joint activities can be achieved through talk as well as embodied means (Balantani and Lázaro, 2021; Clark, 1996; Hausendorf, 2013; Haviland, 2007, 2011).

This paper investigates how musicians achieve to coordinate their actions when practicing their instruments together during a rehearsal session. Specifically, the focus is placed on the use of non-lexical vocalisations in conjunction with the deictic expression “so” or “so_was”/‘like this’ in non-instructive settings.¹ Drawing on the methodological principles of Conversation Analysis (Drew, 2005; Sacks, 1992), the paper addresses the interactional significance of the practice vocalisation + “so” or “so_was” in music rehearsals between friends when deciding on the rhythm and tempo of their next performance.² Focusing on the sequences prior to the performance of a piece, where this practice often occurs, I demonstrate that non-lexical vocalisations, when coupled with the deictic expression “so” or “so_was”, form a multimodal package (Goodwin, 2003; Heath, 1986; Stukenbrock, 2010) utilised by participants in the service of establishing a joint decision on

E-mail address: angeliki.balantani@unil.ch.

¹ To my knowledge, there is no research that investigates the interactional difference between “so” and its variant “so_was”. “so_was”, however, is a contraction of “so” + “etwas” (=something) and is translated in my examples as “something like that”.

² Rhythm contains certain elements, one of which is tempo. By vocalising the rhythm of a piece of music, participants are communicating the tempo as well.

how to perform the next part of a musical piece and sharing the responsibility of that decision. Participants vocalise a rhythm and produce the deictic “so” or “so_was” with rising intonation in the next turn. Produced with a try-marked intonation (Sacks and Schegloff, 1979), the deictic “so” or “so_was” refers retrospectively to the prior embodied action, the suggested rhythm, and projects a confirmation or disconfirmation of that action as a relevant next (Schegloff and Sacks, 1973). By seeking their interlocutors’ approval of the suggested rhythm and tempo, participants orient to their shared deontic authority (Stevanovic and Peräkylä, 2012) with regard to decisions about their joint performance.

Following this brief introduction, section 2 introduces some literature on joint decision making and how deontic authority is constructed and negotiated in such settings. I give a brief review of existing studies on the use of non-lexical vocalisations and the German deictic “so” and describe the gaze and gesture practices associated with the token. In section 3, I provide a note on the data and methodology, followed by the empirical analysis on the use of vocalisations + “so_was” as a multimodal package (section 4). I provide four cases, in two of which the vocalisations are initiating actions proposing a certain rhythm, and in the other two they are responses to a previous suggestion embodying a candidate demonstration. The paper concludes with a discussion of the empirical findings (section 5).

2. Background

2.1. Symmetrical decision making

Joint activities require some form of negotiation and shared decision making between the participants involved. One way of initiating the negotiation process is by suggesting a plan of action that invites the other interlocutors to either accept or reject it (Herder et al., 2018; Kämäräinen et al., 2020; Stevanovic, 2013a; Stevanovic and Peräkylä, 2012; Wasson, 2016). The proposed action plan can be formulated in various linguistic formats, such as interrogative, declarative or imperative, each of which suggests different entitlement with regard to the decision to be made and the relative distribution of deontic rights between the participants (Couper-Kuhlen, 2014; Stevanovic, 2013b). The notion of deontic rights refers to a speaker’s potential to determine other’s future action (Stevanovic, 2013b) and their distribution is locally organised by participants in sequences of interaction; that is, the design of an interlocutor’s utterance may suggest a symmetrical or asymmetrical distribution of deontic rights. Consider the difference between the utterance “I would play that one first” formulated as a declarative and “Should we play that one first?” formulated with interrogative syntax. While in the first utterance, formulated as a stating conditional, the decision about the future action is presented as already established, the second one conveys a request to participate in joint decision-making about the proposed future action (Stevanovic, 2013a). The speaker is asking the recipient for his/her agreement to the proposed action plan; hence their future actions are contingent on the recipient’s approval. Thus, depending on how a suggestion is formulated, speakers can negotiate their deontic rights. They can claim their deontic authority on a matter at hand, challenge their interlocutor’s authority or take a weakened deontic stance and suggest a more symmetrical distribution of deontic rights (Stevanovic, 2013b). While there are many linguistic resources available for interlocutors to construct and negotiate their deontic authority, the actions implicit in the linguistic formats are subtly negotiated in the turn-by-turn sequential unfolding of interaction. Recipients may accept the deontic claims expressed by the speaker by providing a congruent response to the action initiated. For instance, a congruent response to a proposal is an approval of that proposal. However, participants can also challenge and undermine the speakers’ claim to those rights by responding to the first action with incongruent responses (see, for instance, Stevanovic and Peräkylä, 2012 for the *announcement of decisions* as incongruent responses to proposals).

In the current study, I have identified instances where interlocutors produce a non-lexical vocalisation as an embodied demonstration of the rhythm and tempo to be followed and the deictic expression “so” or “so_was” with rising intonation in their next turn. I argue that the vocalisations, in combination with the deictic expression “so_was”, form a multimodal package (Goodwin, 2003; Hayashi, 2005; Heath, 1986; Stukenbrock, 2010), with which participants invite their addressees to approve or disapprove of the suggested tempo and rhythm and establish a “symmetrical distribution of deontic rights” (Stevanovic and Peräkylä, 2012) on determining how to proceed.

2.2. The use of non-lexical vocalisations in social interaction

Non-lexical vocalisations are semantically and syntactically empty syllables that possess no lexical meaning but can be a rich semiotic resource in accomplishing different social actions (Goodwin et al., 2002; Keevallik and Ogden, 2020; Tolins, 2013). These liminal signs (Dingemanse, 2020), combined with gestures and other semiotic resources, are shown to contribute to action formation in their local environments (Goodwin et al., 2002). Non-lexical vocalisations have been examined in different sequential environments, from interactions with persons with aphasia (Goodwin et al., 2002), to enactments in storytellings (Clark, 2016; Fox and Robles, 2010; Goffman, 1978; Golato, 2000; Hsu et al., 2021; Sidnell, 2006; Streeck, 2002) and, more relevant to this study, as an embodied resource in the context of music and dance instructions (Haviland, 2007; Keevallik, 2014, 2015; Tolins, 2013; Weeks, 1996). Weeks (1996) and Haviland (2007) examine vocalisations used in corrective sequences in which the conductor mimics a musical piece. Their occurrence in corrective sequences is also observed by Tolins (2013) whose study focuses exclusively on non-lexical vocalisation in one-to-one music tuitions. He demonstrates that vocalisations are used by instructors in accomplishing two distinct actions, assessing and directing; assessing a piece that was previously played by the student in order to highlight parts of it that need improvement and directing how a piece should be played in future

performances. Vocalisations also feature as part of embodied demonstrations in dance instructions (Keevallik, 2014, 2015) where non-lexical vocalisations are used to accompany the demonstration, thus filling up the vocal-turn space.

This paper will not focus on the use of non-lexical vocalisations per se but will examine their occurrence as part of a multimodal package. The fundamental characteristic of the examples presented in this study is that the vocalisations are not part of an instruction sequence (Haviland, 2007; Keevallik, 2014, 2015; Tolins, 2013; Weeks, 1996).³ In instructional sequences, the vocalisations are part of the instructor's turn and are used to perform specific actions, such as assessments and instruction giving (Tolins, 2013). The participants in my data, however, are friends, who occasionally meet to practice their instruments, their level of competency is more or less the same and, hence, their relative knowledge of the piece they are currently playing is the same. Their goal is not to teach and learn but to learn together, a characteristic that inevitably shapes the sequential environment in which the vocalisations feature and, in turn, the action they perform. The present study also differentiates itself from Keevallik's (2014, 2015) examination of vocalisations. While in bodily-vocal demonstrations vocalisations are merely accompaniments to the embodied demonstration, in the current study vocalisations are the embodied demonstration that coordinated with other concurrent communicative acts (Streeck, 2013, p. 675) or semiotic resources (Goodwin, 2000) form a multimodal package (Goodwin, 2003; Heath, 1986; Stukenbrock, 2010) or a multimodal Gestalt (Mondada, 2014).⁴

2.3. German "so" as a multimodal package

Studies on the deictic use of "so" illustrate that the token functions as a grammatical link between the speaker's gesture or bodily performance and the addressee's gaze practices (Stukenbrock, 2010, 2014, 2015), as it invites the addressee to look at the embodied action of the speaker. Streeck (2002) also documents that "so" as a "flag" is accompanied by speakers' gaze shift to their hands when introducing an iconic gesture in order to alert their addressees' attention to the extralinguistic information that is to be gathered from the gesturing hands. Gaze then plays an important role in modal deictic practices⁵ and attention sharing, not only in the participants' on-line organization of joint attention but also in the establishment of intersubjectivity. While in the aforementioned studies "so" is accompanied by "concurrent" gestures, the deictic reference in this study accompanies "component" depictions. In multimodal packages of "so" with iconic gestures, the token precedes the gesture and therefore aids in introducing the nonverbal behaviour that is imminent. The speaker's gaze direction attracts the addressee's visual attention to the location where this additional meaning can be found (Streeck, 2002). In contrast, the token in this study follows an embodied demonstration that depicts a referent (Clark and Gerrig, 1990) and is anaphoric in nature. Participants may produce some gestures that accompany the embodied demonstration, such as tapping or snapping their fingers to the beat, but these are used complementarily to the vocalisation and do not form part of the demonstration.

The vocalisations are produced either as initiating actions, proposing a tempo and rhythm, or as responsive actions to a previous suggestion, as candidate embodied demonstrations of that lexical suggestion. In other words, the verbal description proposes a type, and the vocalisation is a token that is offered, an indexical representation of the suggestion. While the vocalisations, in this case, function as candidate embodied demonstrations of the rhythm, with the deictic "so_was" participants seek their co-participants' confirmation of their candidate understanding in order to accomplish coordination and cooperation between them in the performance of a musical piece (for a discussion on the German deictic "so" in seeking a co-participant's confirmation of an action, see Balantani, 2022). By seeking confirmation to their prior embodied demonstration, speakers suggest that the proposal to proceed with their demonstrated tempo and rhythm is contingent on the recipient's approval. Addressees, in turn, direct their gaze to the sheet music and follow the suggested demonstration there. Thus, they accept the embodied demonstration of the music as merely a suggestion that needs to be checked and not as something the speaker has authority over to decide on his/her own. The purpose of this study is therefore twofold. First, it seeks to provide an understanding of the deontic issues involved in the use of the multimodal package vocalisation + "so_was". In particular, I describe how participants in non-instructive music sessions make use of this multimodal package in the sequences that precede the performance to orient to their deontic rights with respect to making a decision on how to proceed. Second, by considering the speaker and addressee's eye-gaze in these sequential environments, I hope to contribute to continuing research on the role of eye-gaze in multimodal packages with "so".

3. Methodology and data

This study has been conducted using the methodological principles of Conversation Analysis (Drew, 2005; Sacks, 1992), which in terms of multimodal analysis describes how talk, gesture, gaze, body posture and the physical

³ Only Tolin's (2013) study focuses exclusively on vocalisations. Weeks (1996) and Haviland (2007) examine the interplay of the different semiotic resources that the instructor uses during instruction in orchestra rehearsals, one of which is the use of vocalisations to represent music. Hence, their focus is not the study of vocalisations per se.

⁴ "Demonstrations" are embodied actions that are produced for others to imitate or avoid in the future and differ from "depictions" that are to be interpreted and translated into future embodied performances (Szczeppek Reed, 2021). In light of this distinction, I define the non-lexical vocalisations produced by the speakers in this study as demonstrations of a rhythm to be imitated by themselves and their addressees in their performance of the next piece.

⁵ "Modal deixis" concerns deictic reference to the manner or the quality of an object or an action (Stukenbrock, 2015). In German, modal deixis is expressed by the manner adverbial "so".

surroundings of the participants are jointly used in the performance of social action (Deppermann, 2013; Deppermann and Streeck, 2018; Goodwin, 2017; Kärkkäinen and Thompson, 2018; Keevallik, 2018; Mondada, 2014; Streeck et al., 2011). Data recordings were conducted with the use of mobile eye-tracking glasses (Tobii Pro Glasses 2) worn by the participants and an additional third camera to account for embodied conduct not visible through the eye-tracking. The recordings of the eye-tracking glasses and the third camera were synchronized into one split-screen video and imported into ELAN (2019) for transcription. The verbal transcriptions follow the GAT transcription system developed by Selting et al. (2009) and the embodied conduct was transcribed according to Mondada's (2019) conventions for multimodal transcription (see Appendix).

The use of the eye-tracking glasses in this particular study, permits us to observe whether eye gaze is an important factor in deictic referencing with “so” in the context of music rehearsals, where “hearing” rather than “seeing” is expected to be more relevant. However, as we shall see, the eye-gaze patterns observed in the participants involved indicate that eye gaze is used both by the speaker and the addressee to do specific things. And although gaze fixation is a result of an algorithmic calculation, fixation is not used as a measurement of eye gaze. For this reason, in addition to the eye-tracking glasses, the data was collected using an external camera. Gaze is transcribed using the external camera and by observing the participant's head movement before there is even a fixation. As Goodwin (1994) pointed out, the fact that participants are looking at a common target, for instance, does not necessarily mean that they are “seeing” the same thing. Similarly, a fixation on an object or participant at a certain moment in interaction is not taken as an indicator for gaze direction unless it is made obvious in the interaction as well.

The data analysed is part of a larger corpus (26 h of recordings) on naturally occurring interactions in diverse settings in German and Swiss German collected for a broader research project on deixis and joint attention funded by the Swiss National Science Foundation (SNSF). For the purpose of the current study, I focus on three sets of recordings (2 h and 5 min approximately) that pertain to music rehearsals between friends. The instruments being played are guitars in one of the recordings, violins in another and a saxophone and two clarinets in the third recording. In this paper, I will present examples from the recording of the violin players, that I name “Bratschen_01”, and the saxophone and clarinets, named “Klarinetten_01”. The objective of these meetings is for the participants to practice their instruments by performing a piece of music together. It is not an environment of music instruction where the purpose of the session is to teach and learn respectively. Both participants in the recordings have equal knowledge and level of expertise on the musical piece and their instruments and the aim is to advance their skills by practicing together.

With respect to the overall structural organisation of the interactions, participants are constantly switching between playing their instruments and temporarily pausing the activity to engage in interaction. These pauses take the form of a metatalk about the piece the participants just finished playing and how they are going to proceed. The vocalisations under examination occur late in these “interaction-pauses”, namely just before the interlocutors begin playing their instruments again. This is structurally the place where interlocutors negotiate and agree on the rhythm and tempo of the musical piece they are about to perform. Several studies have focused their research on the transition between discussions about the imminent performance and the performance itself, preparing the start of the latter (for an overview of studies that investigate such transition phases in performing arts, see Ivaldi et al., 2021). In their study on vocal master classes, Szczepek Reed et al. (2013) investigate the negotiations that accompanists and singers undergo when coordinating their restarts after a master's instruction. Participants need to identify and distinguish between two types of directives from the master: *local directives* that require immediate compliance and *restart-relevant directives* that make compliance relevant at a later stage. Schmidt (2018) also looks at the resources deployed in theatrical rehearsals that anticipate an upcoming performance and identifies two practices in the transition spaces, what he calls *projections by arrangements* and *preparations*. These preparatory practices are important parts of the staging process. Similarly, in the present study, before the start of their performance, participants need to achieve a coordination in the different “voices” that each one is performing in. For that they sometimes count in numbers and often they use nonsense syllables, such as, for instance, “tara ta ta” or “pam pam pam”. The collection consists of 5 instances of counting in numbers and 34 instances of non-lexical vocalisations that are either stand-alone or accompanied by the deictic “so” or “so_was”. I will focus on the use of non-lexical vocalisations in combination with the modal deictic “so” or “so_was” in the sequences preceding the playing of the instruments.

4. Empirical analysis: shared decision-making

In what follows, I present data from music rehearsals where participants discuss how they should proceed in the performance of the next piece on their sheet music. The non-lexical vocalisations in the data under consideration are produced by participants as part of a multimodal package that contains the vocalisation as an embodied demonstration of the rhythm and tempo accompanied by the deictic expressions “so” or “so_was”. I will provide four instances from two different data sets that illustrate how the investigated practice is used in the context of practice sessions in deciding what the right tempo and rhythm of a piece is and sharing the responsibilities of that decision. During this “negotiation process”, the non-lexical

vocalisations are produced either as initiating actions, proposing a tempo and rhythm (ex. (3) and (4)), or in response to a previous suggestion, as candidate embodied demonstrations of that lexical suggestion (ex. (1) and (2)). In both cases, the vocalisation is followed by the deictic “so” or “so_was” with rising intonation, which makes a response by the addressee conditionally relevant (Schegloff and Sacks, 1973) and involves him/her in the decision-making activity about the next steps, namely the tempo and rhythm to be followed in the next piece.

In the first excerpt, there are three participants; two are playing the clarinet and one is playing the saxophone and their seating arrangement is side-by-side. They just changed the sheet music on the music stand to a new one to play a different piece. Before beginning to play, they are reading the notes on Jessica’s sheet to find out whether one of the notes is a C or a D.⁶ The excerpt contains a number of attempts at vocalising how to play the piece from different speakers. I will concentrate on the two uses of vocalisations accompanied by the deictic “so” at lines 12 and 32. While both occurrences of “so” form a multimodal package, as the modal deictic reference is finely coordinated with the embodied demonstration, the vocalisation, they are distinct in their interactional function. The “so” at line 09 is used prospectively by the speaker as a “flag” (Streeck, 2002; Stukenbrock, 2010, 2014) that alerts the interlocutors’ auditory attention to the upcoming embodied demonstration in the next turn. At line 34, on the other hand, the speaker uses “so” retrospectively as a confirmation seeking device to her prior embodied behaviour. Therefore, while in both instances the embodied action, the vocalisation, points to the quality of “so”, the practice “so” + vocalisation flags an upcoming vocalisation, without actively seeking any uptake from the addressee, whereas with vocalisation + “so” requests a confirmation to a prior embodied behaviour. Additionally, the stance that the speaker takes towards the piece in each case differs. While in the first instance the speaker takes a more knowing stance, producing the deictic “so” and subsequently demonstrating her own suggestion in an embodied way, in the latter, the stance the speaker is taking is one of uncertainty.

(1) Klarinetten 01 26:51-27:22 “C or D”:

(J: Jürgen, C: Caroline, Je: Jessica)

* for gaze by J

+ for gaze by C

§ embodied conduct by J

& embodied conduct by C

% for gaze by Je

```

01 C aber C würde sinn machen.=
    but c would make sense
02 Je =JA macht sinn,=
    yes makes sense
03 Je =ja <<all> du d[u DU du du DU du du DU du du,> ]=
04 C [( ) c (0.1) c DUR << :->> akkord,>]
    ( ) c (0.1) c major chord
05 Je =ja Eben.
    yes exactly
06 Je das hört sich dann schon GUT an.
    it sounds good then
07 Je ((lacht))
    ((laughter))
08 (1.2)+(0.1)+
    c +.....+
09 C +&al[so %SO; ]&
    so like this
    c +gz at J-->1.15
    c &.....&
    je %gz at C-->1.11
10 Je [((lacht))]
    ((laughter))
11 (0.1)%
    je -->%
12 C &PA (0.2)*pa*(0.2) pa (0.2) pam;=
    c &beats-->1.21
    j *gz at C*
13 J =(anT[Ikenstyle;])

```

⁶ Jessica is sitting in between the two other participants.

```

14 C      [Oder?      ]
      right
15 Je J+A.
      yes
      c      -->+
16 J ta ta tan TA:N tatata TA:N tatata TAN,
17 (0.6)
18 C a[HA:;]
19 Je [D ]A ra ra;
20 (0.2)
21 J <<all> aber& wir KÖNnen es auch erstmal ein bisschen;>;
      but we can first a little
      c      -->§
22 C P[F::      ]
23 Je [((lacht))] ((lacht))
      ((laughter))
24 (0.3)
25 J ((lacht)) [((lacht))]
      ((laughter))
26 Je [((lacht))] ((lacht))
      ((laughter))
27 (0.7)
28 J [ich find das ein BISSchen_ne eigentlich wie_n [(rAffe) ]
      ((lacht))]
      I find it a bit actually like a (snatch) ((laughter))
29 C [HM hm hm (.) hm,
30 Je [((lacht))]
      ((laughter))
31 (1.2)
32 C <<p> TAK +*tak+ tara [TA;>]*
      c      +.....+gz at J and Je-->1.34
      j      *gz at C-----*
33 Je [TA ]<<pp> ta [ta ta] [ta ta;>]
34 C [SO?+ ]
      like this
      c      -->+
35 J [hm_hm, ]

```

Having suggested to play the piece in C major chord (l.04), Caroline proceeds to demonstrate the melody of the piece in that chord in an embodied way (l.12). This focus shift from the tonality to the melody of the piece is marked with the turn-initial reformulation marker “also”/“so” (Deppermann and Helmer, 2013). At line 09 Caroline shifts her gaze to Jürgen and produces “also SO”/“so like this”, the deictic “SO”/“like this” serving as a “flag” for her upcoming embodied action (Streeck, 2002; Stukenbrock, 2010, 2014). Her “flag” attracts Jessica’s gaze but not Jürgen’s, who is looking at his sheet music. Nevertheless, since the embodied action is not a gesture that is to be visually perceived but a vocalisation that requires the addressees’ auditory attention, the absence of Jürgen’s gaze is not problematic. During her vocal demonstration, which is produced in a slow tempo (l.12), Caroline recounts the beats in an embodied way as well by tapping her fingers on her leg. On the second beat, Jürgen directs his gaze to Caroline and, at the end of her demonstration, proposes a different style, namely to play it in “antikenstyle” (l.13).⁷ Caroline’s receipt of Jürgen’s suggestion with the change-of-state token “aHA:;” (l.18) indicates that she has undergone some change in her state of knowledge (Heritage, 1984). Her turn, then, at line 32 “TAK tak tara TA;”, produced in a quicker tempo, is her candidate understanding of the newly proposed way of playing the piece which she follows up with a confirmation seeking “SO?” (l.34). At the end of her second beat, she directs her gaze at her addressees monitoring their attention and Jürgen directs his gaze at her at the same time until the end of her vocalisation. So, we have a mutual monitoring of each other during the demonstration. Since Jürgen proposed this new style, the confirmation seeking “SO?” (l.34) is directed at him. Having produced her candidate understanding of his suggestion, Caroline takes an unknowing stance towards her prior embodied demonstration and, before proceeding to the next step of playing the piece, she needs Jürgen’s approval of her candidate vocalisation. Only after the confirmation is produced at line 35 “hm_hm,” can the participants proceed with their joint performance.

Agreeing on the rhythm and tempo before proceeding with the performance of a piece is important in the context of practice sessions, where participants have to collaborate in order to produce music together. In such contexts, the practice vocalisation + “so_was” proves to be a resource utilised by interlocutors in their attempt to coordinate themselves. To this end, I will examine three more excerpts in which the vocalisation is followed by a try-marked (Sacks and Schegloff, 1979) deictic “SO_was” that demonstrate how this practice is utilised by participants in the service of establishing a joint decision on how to perform the next part of a musical piece and sharing the responsibility of that decision. In excerpt (2), Steven and Lydia play the violin and are standing facing each other. Prior to this sequence, they had trouble with synchronisation in a specific part of the piece they were playing and were trying different “takts” (beats per bar). They eventually agreed that the quicker one sounds better. Having found the right tempo for the previous piece they were playing and before moving on to the next one, the

⁷ By “Antikenstyle” I believe Jürgen refers to *Stile antico*, a music style that was used in the “transitional period between the Renaissance and Baroque eras. The *stile antico* was the style of music based upon Giovanni Palestrina’s polyphonic musical style, used mostly for sacred music, and written around the year 1600.” (<https://www.freemusicdictionary.com/definition/stile-antico/>, accessed 16/02/2021, 16:26).

participants are considering what the right tempo for this new piece could be. The practice, vocalisation + “SO_was”, is used here by Steven as a resource to implicate his co-participant in the decision-making process about the tempo to proceed with. Steven produces a vocalisation at line 15 “pam pam pam PAM,” and with the deictic “SO_was” at line 18 he casts his prior demonstration as a suggestion that requires the recipient’s approval. In this way, the recipient is implicated in making the decision on what tempo to proceed with, as the speaker orients to the fact that he is not fully entitled to make that decision.

(2) Bratschen 01 26:00-26:16 “one quarter”:

(S: Steven, L: Lydia)

* for gaze by L

+ for gaze by S

\$ embodied conduct by L

& embodied conduct by S

```

01  S  ä:hm;
      ehm
02  (1.6)
03  S  <<pp> (tuTU) tu TU tutu TU>
04  L  pf ÄHN[liches tempo;      ]
      pf similar tempo
05  S  [ich hab Irgendwie KEIn]e: keine <<all> erinnerung mehr an
      das stück.>
      I somehow have no no memory anymore of the piece
06  (0.2)
07  L  i[ch AUCh ni]cht.=
      I don't either
08  S  [ähm;      ]
      ehm
09  L  =wir SPIElen==
      we play
10  L  =ja DOCH also;
      yes yes well
11  S  (0.3)
12  L  i[ch würd sagen ÄHNliches tempo;_oder?]
      I would say similar tempo right
13  S  [((flüstert ein Tempo))      ] ((flüstert ein Tempo))
      ((whispers a tempo))
14  (0.2)
15  S  +&pam+[pam pam PA]*M,&
      s +.....+L face-->1.21
      s &beats with the head&
      l          *$ face-->1.17
16  L  [(ist_nur $ein VIER( $ ))]
      it is only one four ( )
      l          $PG w/bow--$
17  (0.1)*
      -->*
18  S  SO_was?
      something like that
19  L  [JA;      ]
      yes
20  S  [<<all> eins VIERTel,>]
      one quarter
21  (0.2)+(0.4)
      s -->+
22  L  (eins) äh VIERTel;
      one eh quarter
23  (0.4)
24  S  hm_[hm;  ]
25  L  [hm_hm, ]

```

The deictic use of “SO_was” at line 18 is anaphoric as it refers to the quality of Steven’s vocalisation of the tempo and his embodied display of understanding at line 15 “pam pam pam PAM.” By vocalising the beats, Steven displays his understanding of the “similar tempo” that his interlocutor previously suggested “ich würd sagen ÄHNliches tempo oder?”/‘I would say similar tempo right’ (L12). His gaze during the vocalisation is on his interlocutor and is kept on Lydia until he receives a confirmation “JA;”/‘yes’ (L19) to his confirmation seeking “SO_was?”/‘something like that’ (L18). Similar to the previous extract, however, his interlocutor’s gaze is on her own sheet music; Lydia does not shift her gaze towards Steven until the end of his vocalising turn. Nevertheless, her confirmation at line 19 demonstrates that, while her visual attention is not on Steven, her auditory attention is. By producing the try-marked “SO_was” (L18) after his candidate understanding of the “similar tempo” suggested by Lydia, Steven takes an unknowing stance and requests his interlocutor’s confirmation to the prior embodied display. Again, what we can see in this excerpt is two response mobilising features, gaze and rising intonation (Stivers and Rossano, 2010), being used by the speaker. The speaker’s gaze is being kept on his interlocutor while demonstrating the tempo and with the use of the try-marked “SO_was”, the recipient is held accountable for responding and is implicated in the decision-making process. The weakening of his deontic stance (Stevanovic, 2013b) with “SO_was” casts his prior demonstration as a suggestion that requires the recipient’s approval. In this way, the recipient is implicated in making the decision on what tempo to proceed with as the speaker orients to the fact that he is not fully entitled to make that decision. After the confirmation at line 19 “JA;”/‘yes’, Steven verbalises his embodied display

"<<all> eins VIERTel,>"/'one quarter' (l.20), thus confirming with Lydia that they share the same understanding of the just completed embodied display of the tempo before they can proceed with the next step, the playing of the piece.

A similar use of the practice vocalisation + "so_was" can be observed in the next extract that comes from the same recording as the previous example. However, the vocalisation in this excerpt is not produced in response to a previously suggested tempo. Instead, the speaker initiates a suggestion of a tempo to follow by vocalising it and produces "SO_was" in his next turn, thus treating his embodied demonstration as contingent on the recipient's approval. Prior to this excerpt, Lydia and Steven decide to move on to play the last part of the piece they are currently playing, the allegro. Before proceeding to play their instruments, they need to see what notes each of them has on their respective sheets so as to coordinate and perform the part correctly.⁸

(3) Bratschen 01 42:36-42:49 "achtelkette":

(S: Steven, L: Lydia)

* for gaze by L

+ for gaze by S

\$ embodied conduct by L

& embodied conduct by S

```
01 L also das alLEgro ist-
    so the allegro is
    l >>gz at sheet-->1.11
02 L ICH hab da eine: (0.2) DURCHgehende achtelkette;
    I have there a continuous chain of eighth notes
03 (0.2)
04 S ich AUCH;
    me too
05 (0.2)
06 L WUNderbar;
    wonderful
07 (0.3)
08 S tarata+tAtataTA ta [tAtata; ]
    s +gz at L-->1.14
09 L [ <<dim> pApapa ] [pAp ] a>;
10 S [SO_was?]
    something like that
11 (0.2)*
    l -->*
12 S [ <<pp> tata ] tAta>
13 L [ *DAS sind achtel; ] *
    these are eighths
    l *gz at S-----*
14 S TA ta ta ta tA+ ta ta [ta tA] ta ta;
    s -->+
15 L [JA; ]
    yes
16 L versUCH ma;
    let us try
17 L mal SEhen wie weit ich komme;
    let us see how far I get
```

The vocalisation of the referent "Durchgehende achtelkette" by Steven at line 08 followed by "SO_was" is treated as a suggestion that needs to be confirmed by his interlocutor. By seeking confirmation to his embodied demonstration, he conveys a weaker deontic stance (Stevanovic, 2013b) to the matter at hand. The proposal to proceed with his suggested tempo is presented as contingent on the recipient's approval. In order to coordinate themselves for the next part they are about to play, they need to be cooperative and determine together how to proceed. And that is what "SO_was" projects in this case; the need for a symmetrical distribution of their rights to establish a joint decision and share the responsibility of that decision. Formulated with an interrogative prosody, and in conjunction with the eye-gaze to his interlocutor, the deictic expression "SO_was" increases the accountability for a response from his recipient (Stivers and Rossano, 2010). However, the confirmation to his embodied demonstration comes at line 15 after an insert sequence where Lydia initiates repair. Lydia's gaze is on her sheet music during the demonstration and is directed at Steven 0.2 s after the deictic expression. The temporal alignment of her vocalisation at line 09, produced in overlap with and matching in rhythm with the last syllables of Steven's vocalisation (Clayton et al., 2005), as well as her uptake at line 13 demonstrate that, despite the lack of eye-gaze, she has been actively listening to the vocalisation. She initiates repair "DAS sind achtel;"/"these are eighths" (l.13) suggesting that the tempo is too fast. This is subsequently repaired by Steven in his next turn "TA ta ta ta tA ta ta ta tA ta ta;" (l.14) in which the vocalisation is produced in a slower tempo. Having established the tempo of the next part, Lydia produces the confirmation token "JA"/"yes" (l.15) and proposes to give it a try (l.16). The confirmation is provided by Lydia only after she has negotiated the right tempo by repairing Steven's prior tempo suggestion and he has corrected it. Thus, we can see how the embodied demonstration that precedes the confirmation seeking "SO_was" is treated by the recipient as merely a suggestion and not a decision that is established.

Let us consider one last excerpt where the practice vocalisation + "so_was" is used in the negotiation of the tempo. Excerpt (4) differs from the excerpts provided so far in that there is no verbal uptake of the participant's vocalisation and subsequent

⁸ The two participants play different "voices", so their sheet music is different.

confirmation seeking “SO_was?”. However, as we shall see, the interlocutors align with the speaker in different ways in overlap with her embodied display. So, what seems to be a “self-confirming” turn at line 14 is in fact a collaboratively produced accomplishment. Jürgen, Caroline and Jessica are about to pick out the next piece they are going to play and Caroline requests if they can play something brisker than the last piece. Jürgen then suggests they play it in Rondo when Caroline enquires about the tempo and subsequently produces a vocalisation as a candidate suggestion at line 10 “papidadidaTA: dadi dididididüdadidandü;”.

(4) Klarinetten 01 17:13-17:26 “what is the tempo like”:

(J: Jürgen, C: Caroline, Je: Jessica)

* for gaze by J

+ for gaze by C

^ for gaze by Je

\$ embodied conduct by J

& embodied conduct by C

£ embodied conduct by Je

```

01  C   wie IST des so:,
      how is it like
02  + (0.5)
      c   +gz at J-->
03  C   TEMPomäfs+sig,
      in terms of tempo
      c   -->+
      je  fairplaying and tapping fingers on keys-->1.10
04  (0.5)
05  C   ich HAB [grad_s ich hab_s grAd nich-^hh          ]
      I have just I have it just
06  J   [(ich weiss es) ehrlich gesagt NICHT mehr;]
      (I do not know that) anymore to tell you the truth
07  (0.2)
08  C   <<p> mhmhm>
09  (1.3) & (0.1)
      c   &beats-->1.14
10  C   papidadida[TA:] dadi dididididüda$didadanf£dü;
      j                                     $snapping with fingers-->1.15
      je                                    -->£
11  J   [ <<pp> hm,>]
12  C   irgend SO_was?
      something like that
13  (0.2)
14  C   ja& +hätte ich je+tzst AUCH $ge*dach*t;£=
      yes I would have also thought so
      c   -->£
      c   +.....+gz at J-->
      j   *.....*gz at C-->1.16
      j   $nods-----$
15  Je =JA: ,=+
      yes
      c   -->+
16  J   =JA.*$
      yes
      j   -->*
      j   -->$

```

The vocalisation in this excerpt comes after a dispreferred response from Jürgen “(ich weiss das) ehrlich gesagt NICHT mehr;”/“(I do not know that) anymore to tell you the truth” (1.06) to Caroline’s enquiry about the tempo of the next piece at lines 01–03 “wie IST es so;”/“how is it like” “TEMPomässig,”/“in terms of tempo”. Jürgen was the one to suggest playing the Rondo so at the end of line 01 Caroline directs her gaze at him and selects him as the next speaker. His non-answer response that he does “not know anymore”, while addressing the relevance of a response, fails to collaborate in progressing the activity further (Stivers and Robinson, 2006). Furthermore, the claim to lack of access to knowledge of the tempo projects the emergence of a joint decision-making sequence about it. In fact, Jessica already begins “airplaying” and tapping her fingers on the keys in the middle of “TEMPomässig”/“in terms of tempo” (1.03) when it is projectable that they are in search of the tempo. After Jürgen’s non-answer, Caroline joins Jessica in the search for the right tempo. She first imitates the melody by humming the notes of her sheet music (1.08) and subsequently produces a vocalisation as a candidate suggestion of the tempo at line 10 “papidadidaTA: dadi dididididüdadidandü;”. Her vocalisation is accompanied by a tapping of her fingers on her knee and slight head movement to the beat. Her interlocutors, Jürgen and Jessica, accompany her vocalisation in an embodied way as well. Jürgen hums silently in overlap and snaps his fingers to the beats towards the end of her vocalising turn as a demonstration of entrainment with Caroline’s rhythm, while Jessica’s finger taps on the keys of her clarinet complement the tempo of Caroline’s vocalisation (Clayton et al., 2005). At line 12 Caroline produces a confirmation seeking “irgend SO_was?”/“something like that” (1.12) to her prior candidate demonstration. Although there is no verbal uptake from her interlocutors, Caroline takes her interlocutors’ finger tapping and snapping during her demonstration to be embodied displays of alignment to her suggestion and after 0.2 s she provides the confirmation herself “ja hätte ich jetzt AUCH gedacht;”/“yes I would have

also thought so' (I.14). The use of "AUCH"/'also' indicates that this is a joint decision as all participants implicitly agreed by aligning with her in an embodied way in the prior turns. At the end of her turn (at "gedacht"), Caroline and Jürgen are mutually monitoring each other and Jürgen nods slightly in agreement before his verbal agreement at line 16 "JA."/ 'yes'.

Similar to the prior examples we have seen, the gaze of the participants during the vocalisation is on their sheet music. The speaker reads the notes and vocalises them and the addressees are following in their own sheet music. So, while their visual attention is on their sheet music, their auditory attention is on the vocalisation. The addressees are actively listening to the tempo performed by the speaker and are ready to confirm the correctness of the proposed tempo when asked to with "SO_was"/'something like that'. "SO_was" then is deictically used by the speaker to refer to the prior embodied demonstration. It is anaphorically used as a confirmation seeking device to Caroline's candidate suggestion of the tempo. It orients to the fact that the demonstration of the tempo is merely a suggestion that requires the recipients' approval, it is an approximate interpretation and not something the speaker has an authority over. This is particularly highlighted in this example as the "SO_was" is preceded by the vagueness marker "irgend", thus indexing uncertainty. The future course of action followed in the next piece is a decision that all three players have to agree on and share responsibility for.

5. Discussion

The paper has investigated a systematic use of non-lexical vocalisations in conjunction with the deictic "so" or "so_was" in achieving coordination in a joint activity, namely a music rehearsal. The study has shown that in music rehearsals, such as the one examined in this study, participants use the practice vocalisation + "so" or "so_was" to coordinate their actions and agree on how they are going to proceed, what rhythm they are going to follow or how quick the tempo is going to be. By examining the sequences that precede the performance of a piece, I show that non-lexical vocalisations are produced either as initiating actions, proposing a tempo and rhythm (ex. (3) and (4)), or as responsive actions to a previous suggestion, as candidate embodied demonstrations of that lexical suggestion (ex. (1) and (2)). In both cases, the vocalisation is followed by the deictic "so" or "so_was" with rising intonation, which establishes a strong expectation of a confirmation from the addressee (Schegloff and Sacks, 1973). While the turn design of vocalisation + "so" or "so_was" with rising intonation marks it as seeking response, the interactional function of this practice is to establish a "symmetrical distribution of deontic rights" (Stevanovic and Peräkylä, 2012) on determining how to proceed. By seeking confirmation to their prior embodied demonstration, speakers orient to a shared deontic authority with respect to the future plan of action (Stevanovic and Peräkylä, 2012). The proposal to proceed with their demonstrated tempo is contingent on the recipient's approval and so the addressee is involved in the decision-making process. Additionally, in all of the extracts we have seen, the speakers' gaze is on the addressee during the vocalisation. Holding their gaze on their interlocutors beyond the completion of their demonstration and producing "so" or "so_was" with rising intonation in the next turn establishes a strong expectation for an affirmative response. In line with Stivers and Rossano's (2010) earlier work on response mobilising features, the speakers of the vocalising turns use gaze and prosody as resources in mobilising a response and involving the addressee in the decision-making process. Addressees, in turn, direct their gaze to the sheet music and follow the suggested demonstration there. Thus, they accept the embodied demonstration of the music as merely a suggestion and acquiesce to maintain the symmetrical distribution of deontic rights.

Furthermore, the study has focused on the use of vocalisations as part of a multimodal package in a "relaxed" environment where participants are friends whose goal is to practice together and learn from each other. In music instructions (Tolins, 2013) and orchestra rehearsals (Haviland, 2007; Weeks, 1996) where the target is to improve performance and musicality overall, vocalisations are used in corrective sequences to accomplish actions such as assessments of performances and directions on how to improve one's performance. In practice sessions, on the contrary, vocalisations combined with the deictic "so" or "so_was" are used to accomplish coordination and cooperation between participants in the performance of a musical piece. The vocalisations are produced as suggestions or candidate understandings that require the addressee's approval and therefore establish a symmetrical distribution of deontic rights with regard to decision-making.

With respect to the deictic use of "so", the investigated practice has shown a different type of multimodal package from the ones analysed so far. In contrast to the multimodal packages with "so" described in the literature, where "so" is accompanied by "concurrent" gestures (Balantani, 2021, 2022; Balantani and Lázaro, 2021; Streeck, 2002; Stukenbrock, 2010, 2014, 2015), the deictic reference in this study accompanies "component" depictions. The embodied action in these instances is a vocalisation, an auditory cue. By vocalising the tempo and rhythm speakers iconically represent its quality, how the piece should sound. The deictic "so_was" then does not point to the gaze summoning quality of "so" (Stukenbrock, 2010) but to the speakers' shared auditory attention. And although addressees' gaze is not on the speaker during his/her vocalisation, their subsequent agreement demonstrates that their auditory attention is on the speaker's audible demonstration. Hence, by withdrawing their gaze, addressees are in fact doing being a good listener and display their attentiveness to their interlocutor's embodied demonstration.

What is more, in line with Streeck's (2002) observation that the use of "so" with bodily enactments marks a change in "footing" (Goffman, 1981), the vocalisations in this study mark a change in footing from talk to embodiment of the rhythm. The vocalisations act as mimetic representations of the rhythm to be followed with the musical instrument and are produced with the same modality as talk, namely the voice. Similar to enactments described by Streeck (2002) where the speaker pretends to occupy another body and animate certain features of it, the speakers in the instances examined in this study use their voices to animate a musical instrument. Participants over the course of the sequence shift their footing between the "talking" voice and the "animated" voice that is used for the demonstration. The "so" that is produced after the vocalisation marks the end of the "animated" voice and the shift back to the "talking" voice.

In sum, the paper has described a practice that interlocutors make use of in coordinating their actions in joint activities and establish a collaborative decision with respect to their future plans of action (Stevanovic, 2013b). The practice described is the use of non-lexical vocalisations in combinations with the deictic “so_was”. One possible line of future inquiry is to explore the relationship between vocalisations with and without the tag and do a comparative analysis. It would be interesting to see what the absence of a tag in the same sequential environment does in terms of action formation. Future research could build on the results presented here and investigate the role of vocalisations as such in this context.

Funding

This research was funded by the Swiss National Science Foundation (SNSF) as part of the project 179108 “Deixis and Joint Attention: Vision in Interaction” (DeJA-VI).

Declaration of competing interest

The author declares no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Acknowledgments

I would like to thank the participants in my study, our data session group at the University of Lausanne and those at the University of Athens and Neuchatel for inviting me to present the data in one of their data sessions as well as the University of Freiburg for organising the workshop “2. Arbeitstreffen zum Mobilen Eye-Tracking in der Interaktionslinguistik” where this paper was presented. I would also like to thank the members of my project team for their help with data collection and preparation and the Principal Investigator of the project, Prof. Anja Stukenbrock. A special thanks is also dedicated to the two anonymous reviewers for their time and thoughtful comments and efforts towards improving this manuscript. Finally, I would like to thank Petra and Mikey for their help with the music terminology.

Appendix A

GAT 2 transcription conventions (GAT2, Selting et al., 2009; for the English translation see Couper-Kuhlen and Barth-Weingarten, 2011).

Sequential structure

[]	Overlap and simultaneous talk
=	Fast, immediate continuation with a new turn or segment (latching)

In- and outbreaths

° h / h̃	In-/outbreaths of appr. 0.2–0.5 s duration
° hh / hh̃	In-/outbreaths of appr. 0.5–0.8 s duration
° hhh / hhh̃	In-/outbreaths of appr. 0.8–1.0 s duration

Pauses

(.)	Micro pause, estimated, up to 0.2 s duration appr.
(0.5)/(2.0)	Measured pause of appr. 0.5 / 2.0 sec. duration (to tenth of a second)

Other segmental conventions

and_uh	Cliticizations within units
uh, uhm, etc.	Hesitation markers, so-called “filled pauses”
:	Lengthening, by about 0.2-0.5 sec.
::	Lengthening, by about 0.5-0.8 sec.
:::	Lengthening, by about 0.8-1.0 sec.
?	Cut-off by glottal closure

Laughter and crying

haha hehe hihi	Syllabic laughter
((laughs)) ((cries))	Description of laughter and crying
<<laughing> >	Laughter particles accompanying speech with indication of scope
<<:-)> so>	Smile voice

Continuers

hm, yes, no, yeah	Monosyllabic tokens
hm_hm, ye_es, no_o	Bi-syllabic tokens
?hm?hm	With glottal closure, often negating

Accentuation

SYllable	Focus accent
sYllable	Secondary accent
!SYllable	Extra strong accent

Final pitch movements of intonation phrases

?	Rising to high
,	Rising to mid
--	Level
;	Falling to mid
.	Falling to low
<<surprised>>	Interpretive comment with indication of scope
<<f>>	Forte, loud
<<ff>>	Fortissimo, very loud
<<p>>	Piano, soft
<<pp>>	Pianissimo, very soft
<<all>>	Allegro, fast
<<len>>	Lento, slow
<cresc>>	Crescendo, increasingly louder
<<dim>>	Diminuendo, increasingly softer
<<acc>>	Accelerando, increasingly faster

Other conventions

(xxx), (xxx xxx)	One or two unintelligible syllables
(may i)	Assumed wording
((. . .))	Omission in transcript

Appendix B

Embodied actions are transcribed according to the following conventions developed by Lorenza Mondada (see [Mondada 2018](#) for a conceptual discussion).

<https://www.lorenzamondada.net/multimodal-transcription>

* *	Descriptions of embodied actions are delimited between
+ +	two identical symbols (one symbol per participant and per type of action)
Δ Δ	that are synchronized with correspondent stretches of talk or time indications.
*--->	The action described continues across subsequent lines
---->*	until the same symbol is reached.
>>	The action described begins before the excerpt's beginning.
---->>	The action described continues after the excerpt's end.
.....	Action's preparation.
----	Action's apex is reached and maintained.
////	Action's retraction.
ric	Participant doing the embodied action is identified in small caps in the margin.
fig	The exact moment at which a screen shot has been taken
#	is indicated with a sign (#) showing its position within the turn/a time measure.

References

- Balantani, A., 2021. Reference construction in interaction: the case of type-indicative “so”. *J. Pragmat.* 181, 241–258 <https://doi.org/10.1016/j.pragma.2021.05.024>.
- Balantani, A., 2022. Coordinating individual actions in joint activities: the case of the German deictic so. *Lingua*. <https://doi.org/10.1016/j.lingua.2022.103402>. In press.
- Balantani, A., Lázaro, S., 2021. Joint attention and reference construction: the role of pointing and “so”. *Lang. Commun.* 79 (6), 33–52 <https://doi.org/10.1016/j.langcom.2021.04.002>.
- Clark, H.H., 1996. *Using Language*. Cambridge University Press, Cambridge.
- Clark, H.H., 2016. Depicting as a method of communication. *Psychol. Rev.* 123 (3), 324–347. <https://doi.org/10.1037/rev0000026>.
- Clark, H.H., Gerrig, R.J., 1990. Quotations as demonstrations. *Language* 66 (4), 764–805. <https://doi.org/10.2307/414729>.
- Clayton, M., Sager, R., Will, U., 2005. In time with the music: the concept of entrainment and its significance for ethnomusicology. *Eur. Meet. Ethnomusicol.* 11, 3–142.

- Couper-Kuhlen, E., 2014. What does grammar tell us about action? *Pragmatics* 24 (3), 623–647. <https://doi.org/10.1075/prag.24.3.08cou>.
- Couper-Kuhlen, E., Barth-Weingarten, D., 2011. A system for transcribing talk-in-interaction, translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 12, 1–51.
- Deppermann, A., 2013. Multimodal interaction from a conversation analytic perspective. *J. Pragmat.* 46 (1), 1–7. <https://doi.org/10.1016/j.pragma.2012.11.014>.
- Deppermann, A., Helmer, H., 2013. Zur Grammatik des Verstehens im Gespräch: Inferenzen anzeigen und Handlungskonsequenzen ziehen mit also und dann. *Z. Sprachwiss.* 32 (1), 1–39. <https://doi.org/10.1515/zfs-2013-0001>.
- Deppermann, A., Streeck, J., 2018. Time in embodied interaction. In: *Synchronicity and Sequentiality of Multimodal Resources*. Benjamins, Amsterdam. <https://doi.org/10.1075/pbns.293>.
- Dingemanse, M., 2020. Between sound and speech: liminal signs in interaction. *Res. Lang. Soc. Interact.* 53 (1), 188–196. <https://doi.org/10.1080/08351813.2020.1712967>.
- Drew, P., 2005. Conversation analysis. In: Fitch, K.L., Sanders, R.E. (Eds.), *Handbook of Language and Social Interaction*. Lawrence Erlbaum, New Jersey, pp. 71–102.
- ELAN, 2019, 5.9. <https://tla.mpi.nl/tools/tla-tools/elan/>. (Accessed 16 July 2021).
- Fox, B., Robles, J., 2010. It's like mmm: enactments with it's like. *Discourse Stud.* 12 (6), 715–738. <https://doi.org/10.1177/1461445610381862>.
- Free Music Dictionary (n.d.). *Stile Antico*. In *freemusicdictionary.com*. Retrieved February 16, 2021, from <https://www.freemusicdictionary.com/definition/stile-antico/>.
- Goffman, E., 1978. Response cries. *Language* 54, 787–815. <https://doi.org/10.2307/413235>.
- Goffman, E., 1981. *Forms of Talk*. University of Pennsylvania Press, Philadelphia, PA.
- Golato, A., 2000. An Innovative German Quotative for Reporting on Embodied Actions: und ich so/und er so “and I'm Like/and He's like. *J. Pragmat.* 32, 29–54. [https://doi.org/10.1016/S0378-2166\(99\)00030-2](https://doi.org/10.1016/S0378-2166(99)00030-2).
- Goodwin, C., 1994. Professional vision. *Am Anthropol* 96 (3), 606–633. <https://doi.org/10.1525/aa.1994.96.3.02a00100>.
- Goodwin, C., 2000. Action and embodiment within situated human interaction. *J. Pragmat.* 32, 1489–1522. [https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X).
- Goodwin, C., 2003. Pointing as situated practice. In: Kita, S. (Ed.), *Pointing, Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum, Mahwah, NJ, pp. 217–241. <https://doi.org/10.4324/9781410607744>.
- Goodwin, C., 2017. *Co-Operative Action*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781139016735>.
- Goodwin, C., Goodwin, M.H., Olsher, D., 2002. Producing sense with nonsense syllables: turn and sequence in conversations with a man with severe aphasia. In: Ford, C.E., Fox, B., Thompson, S.A. (Eds.), *The Language of Turn and Sequence*. Oxford University Press, New York, pp. 56–92. <https://doi.org/10.4324/9781003060994-27>.
- Hausendorf, H., 2013. On the interactive achievement of space – and its possible meanings. In: Auer, P., Hilpert, M., Stukenbrock, A., Szmrecsanyi, B. (Eds.), *Space in Language and Linguistics: Geographical, Interactional and Cognitive Perspectives*. De Gruyter, Berlin, pp. 276–303. <https://doi.org/10.1515/9783110312027.276>.
- Haviland, J., 2007. Master speakers, master gesturers: a string quartet master class. In: Duncan, S., Cassell, J., Levy, E. (Eds.), *Gesture And the Dynamic Dimension of Language: Essays in Honor of David McNeill*. John Benjamins, Amsterdam, pp. 147–172. <https://doi.org/10.1075/gs.1.16hav>.
- Haviland, J., 2011. Musical spaces. In: Goodwin, C., Streeck, J., LeBaron, C. (Eds.), *Multimodality and Human Activity: Research on Human Behavior, Action, and Communication*. Cambridge University Press, New York, pp. 289–304.
- Hayashi, M., 2005. Joint turn construction through language and the body: notes on embodiment in coordinated participation in situated activities. *Semiotica* 2005 (156), 21–53. <https://doi.org/10.1515/semi.2005.2005.156.21>.
- Heath, C., 1986. *Body Movement and Speech in Medical Interaction*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511628221>.
- Herder, A., Berenst, J., de Glopper, K., Kooze, T., 2018. Nature and function of proposals in collaborative writing of primary school students. *Ling. Educ.* 46 (1), 1–11. <https://doi.org/10.1016/j.linged.2018.04.005>.
- Heritage, J., 1984. A change-of-state token and aspects of its sequential placement. In: Atkinson, J.M., Heritage, J. (Eds.), *Structures of Social Actions: Studies in Conversation Analysis*. Cambridge University Press, Newcastle, pp. 299–345. <https://doi.org/10.1017/CBO9780511665868.020>.
- Hsu, H.-C., Brône, G., Feyaerts, K., 2021. When gesture “takes over”: speech-embedded nonverbal depictions in multimodal interaction. *Front. Psychol.* 11, 552533. <https://doi.org/10.3389/fpsyg.2020.552533>.
- Ivaldi, A., et al., 2021. Learning to perform: a conversation analytic systematic review of learning and teaching practices in performing arts lesson interactions. *Learn. Cult. Soc. Interact.* 28, 1–17. <https://doi.org/10.1016/j.lcsi.2020.100459>.
- Kämäräinen, A., Eronen, L., Björn, P.M., Kärnä, E., 2020. Initiation and decision-making of joint activities within peer interaction in student-centred mathematics lessons. *Classroom Discourse*. <https://doi.org/10.1080/19463014.2020.1744457>.
- Kärkkäinen, E., Thompson, S.A., 2018. Language and bodily resources: ‘Response packages’ in response to polar questions in English. *J. Pragmat.* 123, 220–238. <https://doi.org/10.1016/j.pragma.2017.05.003>.
- Keevallik, L., 2014. Turn organisation and bodily-vocal demonstrations. *J. Pragmat.* 65, 103–120. <https://doi.org/10.1016/j.pragma.2014.01.008>.
- Keevallik, L., 2015. Coordinating the temporalities of talk and dance. In: Deppermann, A., Günther, S. (Eds.), *Temporalities in Interaction*. John Benjamins, Amsterdam, pp. 309–336. <https://doi.org/10.1075/slsi.27>.
- Keevallik, L., 2018. The temporal organization of conversation while mucking out a sheep stable. In: Deppermann, A., Streeck, J. (Eds.), *Modalities and Temporalities: Convergences and Divergences of Bodily Resources in Interaction*. Benjamins, Amsterdam. <https://doi.org/10.1075/pbns.293.03kee>.
- Keevallik, L., Ogden, R., 2020. Sounds on the margins of language at the heart of interaction. *Res. Lang. Soc. Interact.* 53 (1), 1–18. <https://doi.org/10.1080/08351813.2020.1712961>.
- Mondada, L., 2014. The local constitution of multimodal resources for social interaction. *J. Pragmat.* 65, 137–156. <https://doi.org/10.1016/j.pragma.2014.04.004>.
- Mondada, L., 2018. Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51 (1), 85–106. <https://doi.org/10.1080/08351813.2018.1413878>.
- Mondada, L., 2019. *Transcription Conventions*. Retrieved September 16, 2020, from <https://www.lorenzamondada.net/multimodal-transcription>.
- Sacks, H., 1992. In: Jefferson, G. (Ed.), *Lectures on Conversation*, I+II. Blackwell, Oxford.
- Sacks, H., Schegloff, E.A., 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In: Psathas, G. (Ed.), *Everyday Language: Studies in Ethnomethodology*. Irvington Publishers, New York, pp. 15–21.
- Schegloff, E.A., Sacks, H., 1973. Opening up closings. *Semiotica* 8, 289–327. <https://doi.org/10.1515/semi.1973.8.4.289>.
- Schmidt, A., 2018. Prefiguring the future: projections and preparations within theatrical rehearsal. In: Deppermann, A., Streeck, J. (Eds.), *Time in Embodied Interaction. Synchronicity and Sequentiality of Multimodal Resources*. John Benjamins, Amsterdam/Philadelphia, pp. 231–260. <https://doi.org/10.1075/pbns.293.07sch>.
- Selting, M., et al., 2009. *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*. *Gesprächsforschung - Online-Zeitschrift zur Verbalen Interaktion* 10, 353–402.
- Sidnell, J., 2006. Coordinating gesture, talk and gaze in reenactments. *Res. Lang. Soc. Interact.* 39 (4), 377–409. https://doi.org/10.1207/s15327973rsi3904_2.
- Stevanovic, M., 2013a. Constructing a proposal as a thought: a way to manage problems in the initiation of joint decision-making in Finnish workplace interaction. *Pragmatics* 23 (3), 519–544. <https://doi.org/10.1075/prag.23.3.07ste>.
- Stevanovic, M., 2013b. *Deontic Rights in Interaction: A Conversation Analytic Study on Authority and Cooperation*, 10 2013. Publications of the Department of Social Research. Doctoral dissertation, Department of Social Research, University of Helsinki.
- Stevanovic, M., Peräkylä, A., 2012. Deontic authority in interaction: the right to announce, propose, and decide. *Res. Lang. Soc. Interact.* 45 (3), 297–321. <https://doi.org/10.1080/08351813.2012.699260>.

- Stivers, T., Robinson, J.D., 2006. A preference for progressivity in interaction. *Lang. Soc.* 35, 367–392. <https://doi.org/10.1017/S0047404506060179>.
- Stivers, T., Rossano, F., 2010. Mobilizing response. *Res. Lang. Soc. Interact.* 43 (1), 3–31. <https://doi.org/10.1080/08351810903471258>.
- Streeck, J., 2002. Grammars, words, and embodied meanings: on the uses and evolution of so and like. *J. Commun.* 52, 581–596. <https://doi.org/10.1111/j.1460-2466.2002.tb02563.x>.
- Streeck, J., 2013. Praxeology of gesture. In: Müller, C., et al. (Eds.), *Body-Language-Communication: an International Handbook on Multimodality in Human Interaction*. De Gruyter Mouton, Berlin/Boston, pp. 674–688. <https://doi.org/10.1515/9783110261318.674>.
- Streeck, J., Goodwin, C., LeBaron, C., 2011. *Embodied Interaction in the Material World*. Cambridge University Press, Cambridge.
- Stukenbrock, A., 2010. Überlegungen zu einem Multimodalen Verständnis der Gesprochenen Sprache am Beispiel Deiktischer Verwendungsweisen des Ausdrucks "so". In: Dittmar, N., Bahlo, N. (Eds.), *Beschreibungen für Gesprochenes Deutsch auf dem Prüfstand. Analysen und Perspektiven*, pp. 165–193. Frankfurt/M.: Lang.
- Stukenbrock, A., 2014. Take the words out of my mouth: verbal instructions as embodied practices. *J. Pragmat.* 65, 80–102. <https://doi.org/10.1016/j.pragma.2013.08.017>.
- Stukenbrock, A., 2015. *Deixis in der face-to-face-Interaktion*. de Gruyter, Berlin, Boston. <https://doi.org/10.1515/9783110307436>.
- Szczepek Reed, B., 2021. *Depictions and demonstrations: two methods of communication*. In: Paper presented at the 22. Arbeitstagung zur Gesprächsforschung: Bedeutung in der Interaktion. Mannheim.
- Szczepek Reed, B., et al., 2013. NOW or NOT NOW: coordinating restarts in the pursuit of learnables in vocal master class. *Res. Lang. Soc. Interact.* 46 (1), 22–46. <https://doi.org/10.1080/08351813.2013.753714>.
- Tolins, J., 2013. Assessment and direction through nonlexical vocalizations in music instruction. *Res. Lang. Soc. Interact.* 46 (1), 47–64. <https://doi.org/10.1080/08351813.2013.753721>.
- Wasson, C., 2016. Integrating conversation analysis and issue framing to illuminate collaborative decision-making activities. *Discourse Commun.* 10 (4), 378–411. <https://doi.org/10.1177/1750481316638153>.
- Weeks, P., 1996. A rehearsal of a beethoven passage: an analysis of correction talk. *Res. Lang. Soc. Interact.* 29, 247–290. https://doi.org/10.1207/s15327973rlsi2903_3.

Angeliki Balantani is a postdoctoral research fellow on the SNSF-funded project “Deixis and Joint Attention: Vision in Interaction (DeJA-VI)” where she investigates the role of gesture and gaze in the use of the deictic expression “SO” with the help of mobile eye-tracking glasses. She recently completed her PhD at the University of Essex under the supervision of Dr. Rebecca Clift where she investigated responses to informings in Greek talk-in-interaction.