

Research Article

Hartmut Stoeckl and Monika Messner*

Tam pam pam pam and mi – fa – sol: constituting musical instructions through multimodal interaction in orchestra rehearsals

<https://doi.org/10.1515/mc-2021-0003>

Received January 23, 2021; accepted September 9, 2021; published online October 11, 2021

Abstract: Using multimodal (inter)action/conversation analysis, the present contribution inventories the repertoire of higher-level actions that constitute musical instruction in orchestra rehearsals. The study describes the modal complexity of the instructional actions as built from a varied combination of speech, gesture, gaze, vocalizing and body posture/movement. A high modal intensity of speech and vocalizing is explained with recourse to their contextually useful modal reaches. While some modes, like vocalizing and body posture appear to be action-specific, others turn out to be pervasive default modes. Besides modal intensity, the study also attends to the transitioning between higher-level actions through gaze and the role of the score as frozen action. The analyses help demystify orchestra rehearsals as a special type of professional communicative interaction, which builds on a rich multimodal texture motivated by recurring instructional functions. The methodological rationale demonstrated will be suited to exploring the social variation of instructional interaction in orchestra rehearsals.

Keywords: conversation analysis; higher-level actions; modal complexity; modal intensity; multimodal (inter)action analysis; musical instructions; orchestra rehearsals

1 Introduction

Classical musicology and conversation analysis have looked at orchestra rehearsals mainly as readings of scores and comments on a musical text. Only few approaches have directed attention to the interactional nature of orchestra rehearsals. Professional practice shows that conductors seeking to build a model musical performance engage in complex multimodal interaction, employing not just speech but also gesture, body posture/movement, gaze and imitational singing/vocalizing. The present paper is a first attempt to model the ways in which conductors marshal these different semiotic modes and combine them strategically in order to realize both individual higher-level actions and larger instructional sequences. Owing to their multimodal richness, instructions in orchestra rehearsals are an ideal testing ground for interaction analysis, which we aim to explore, seeking to fill a gap in the multimodal study of institutional/professional communication. On closer inspection, interactional sequences that seem idiosyncratic and creative turn out to reveal a regular, functionally motivated multimodal texture. In this, different semiotic modes take over specific interactional tasks, depending on their respective modal strengths. Rather than aggregate much data, the paper is based on an

*Corresponding author: **Monika Messner**, Institute for Romance Studies, University of Innsbruck, Innsbruck, Austria,

E-mail: monika.messner@uibk.ac.at. <https://orcid.org/0000-0003-0028-2370>

Hartmut Stoeckl, Department of English & American Studies/English and Applied Linguistics, University of Salzburg, Salzburg, Austria, E-mail: hartmut.stoeckl@sbg.ac.at. <https://orcid.org/0000-0002-4517-4464>

analysis of one representative sample drawn from a larger corpus of rehearsal transcripts within the theoretical frameworks of multimodal (inter)action analysis (Norris 2016; Pirini 2014) and multimodal conversation analysis (Deppermann 2013; Mondada 2013, 2016).

Orchestra rehearsals can safely be regarded as a type or genre of communicative activity whose mediational means (i.e., speech plus other modes) are shaped by the dominant function of instructing an orchestra to perform a piece of music in the way laid out in the score and intended by the conductor. This functionality requires the conductor to describe or demonstrate to the musicians the preferred character of musical sound. If we follow Bateman et al.'s argument that modes inherently emerge from or are afforded by material/medial preconditions (cf. Bateman et al. 2017: 101–110, 217–221), the orchestra rehearsal may be labeled as a temporal and unscripted medial-material situation (Bateman et al. 2017: 238–248). This means that face-to-face interaction evolves in time without following a pre-defined design. Add to this the notion that the interaction takes place in a neatly structured space (cf. classic seating arrangement of orchestra sections relative to conductor) and the fact that everyone is immersed and can get involved in the interaction, you arrive at a type of communication that combines dynamic face-to-face interaction with strategic uses of spatial aspects of the situation. Speech, gesture, gaze, body posture/movement and vocalization (singing) then are modes that naturally follow or emerge from this medial set-up.

The interactive practice of orchestra rehearsals is multimodal in essentially two ways: first, the score is 'translated' (cf. 'transduction', Kress 2010: 124) into musical sound and becomes the object of interpretation and interaction between conductor and musicians. Second, this interaction about musical qualities and their performative enactment relies on the cohesive and coherent combination of different semiotic modes to realize various communicative actions, all instrumental in achieving the goal of instructing.

The paper aims at demonstrating on orchestra rehearsals a method of analysis primarily inspired by multimodal (inter)action analysis. This allows insights into a professional and complex communicative practice, which corroborate evidence of the ways in which modes differ in their 'reach' and combine to produce what seem to us generic multimodal patterns typical of the orchestra rehearsal. In addition to applying and fine-tuning a method for the study of a largely unexplored interaction type, we will show the regular multimodal texture of rehearsal instructions, which is governed by the interactional functions necessary and the general logic of the modes involved. In this fashion, the paper contributes to demystifying an apparently erratic professional practice of multimodal communication. It also provides a tailor-made methodological rationale for exploring the social variation of this interaction type that may be due to different conductors, varying language repertoires available and differing degrees of interactional familiarity.

2 Orchestra rehearsals as multimodal (inter)action settings

In this section, we briefly review some of the previous work on rehearsal interaction and available methods for studying it. Generally, orchestra rehearsals, as institutional settings, comprise two actors (social roles): the conductor who does all the instructional talking/conducting and the musicians who play/perform the music and respond to the conductor's instructions. The right to speak is thus (asymmetrically) distributed, favoring the conductor: he/she instructs the musicians verbally and in an embodied way. Such *embodiments* or *embodied demonstrations* are perceived, according to Goffman (1974), as simulations or performances, through which conductors *show* the musicians how to perform a certain passage in the score. This means that the instructional content is not described verbally, but demonstrated (cf. 'demonstrating', Clark and Gerrig 1990) or depicted (cf. 'depicting', Clark 2016) with the help of other modes such as singing or gesture. Conversational asymmetry is intensified by the fact that conductors have access to the score, whereas musicians work from parts.

The musicians *answer/react* by playing their instruments accordingly. The conductor is the "authorized correction-agent" (Weeks 1996: 253) who has the right to interrupt the music at every conceivable point during the rehearsal and to correct/instruct during these interruptions. For this purpose, conductors use a strategic mix of different semiotic modes: e.g., talk, gestures, facial expressions, gaze, singing, proxemics etc. In doing

so, their instructional activities are determined by the score. The score has prescriptive character: it instructs the musicians what, how and when to play, and it forms the basis of the conductor's directing, demonstrating, explaining and describing.

Instructional interaction has been studied in a variety of different musical settings and from various perspectives (Ivaldi 2016; Keevallik 2010, 2015; Nishizaka 2006; Stevanovic and Kuusisto 2018; Szczepek Reed et al. 2013; Tolins 2013). The essential focus of such studies is to understand and describe the ways in which musical teachers perform instructional actions through a combination of talk and embodied action in order to demonstrate the intended sonic interpretation of the music.

In musicology studies, the communication between conductor and orchestra musicians has been explored within the context of rehearsal strategies as well as in relation to leadership and authority (cf. Faulkner 1973; Price and Byo 2002; Biasutti 2012; Adenot 2015). These studies focus on verbal communication and feedback, but they also shed light on a number of non-verbal issues, such as eye contact, facial expression, gestures, singing and “expressive signs” (Faulkner 1973: 150). Their major result is highlighting the importance and effectiveness of non-verbal communication in orchestra rehearsals as Biasutti (2012: 64) argues: “Much more important than words are contact, gesture and eye contact ... language without words”. To some even, the extensive use of talk by the conductor is related to incompetence. Adenot (2015: 8) for instance says: “[m]usicians seem always circumspect with a conductor who is over-verbal in explaining what he wants to do: when, in their view, a conductor talks too much, they suspect he is in some way incompetent”. Price and Byo (2002: 346) characterize the provision of information as happening in a “multimodal (sensory) fashion”.

Recent ethnomethodological and anthropological studies have analyzed the interplay of verbal, vocal and visual semiotic modes in musical settings, such as rehearsals and master classes (cf. Weeks 1996; Haviland 2007). In his study of corrective sequences in orchestra rehearsals, Weeks (1996) identifies different techniques for correcting. Verbal instructions are employed to explain the desired musical effect; illustrative expressions such as singing, chanting and counting with fingers are utilized to embody the music; and through contrast pairs, the conductor juxtaposes a correct/desired way of performing the music and an incorrect/undesired performance. Haviland (2007) shows how in a master class for string quartet, professional music teachers alternate between verbal comments and embodied demonstrations to correct the students' performances. Such embodied demonstrations are realized through a number of different semiotic modes and their combination, including also pantomime (positions of the hands, movements of the arms and manipulation of the bow). These aim at conveying an idea of how a musical passage should be played on a string instrument. Haviland (2007: 170–171) emphasizes that in musical masterclasses bodies “become at once vehicles of performance and meta-performance, means for making music and for communicating about music”.

In a recent conversation analytic study by Veronesi (2014) on conductors' corrections, the “lamination” (Veronesi 2014: 479), i.e., layering of different semiotic modes (cf. Goodwin 2013: 12) and the presence of “multimodal unit[s]” (Veronesi 2014: 480) are observed (cf. also Keevallik 2010: 423). These function to fuse talk and embodiment so as to illustrate the same concept in parallel. The conductor imitates certain instruments (such as percussion) and employs demonstrational singing “as a key resource to locate the correctable and offer a model for correction” (Veronesi 2014: 485). He also employs iconic gestures to exemplify musical characteristics, such as the ‘open hand supine’ (Kendon 2004) for the length of a note (cf. also Bräm and Boyes Bräm 2004). Imitative and iconic gestures as well as sung demonstrations seem very central in a conductor's instructional activities during rehearsals.

Multimodal interaction has been studied in two major, essentially compatible methodological frameworks: multimodal (inter)action analysis (MIA) (Norris 2004, 2011) and multimodally expanded conversation analysis (MCA) (Deppermann 2013, 2018; Mondada 2013, 2016). While both focus primarily on the identification and modal constitution of actions, their methods differ slightly in the approaches to transcription and in the take on how multimodal discourse is generated. MCA uses score transcription enriched by notations for non-verbal phenomena; MIA employs differing conventions with the aim to effectively render the non-verbal modes. Whereas MIA ascribes actions to mode combinations rather categorically, MCA regards actions as dynamically emerging from “multimodal practices” (Deppermann et al. 2016: 13; Mondada 2016: 362). In terms of the relations between modes, MCA departs from absolute mode equality in a holistic interplay; MIA, by

contrast, takes different and shifting “modal intensities” as an inherent given in multimodal interaction. Finally, while discourse is seen to primarily result from a chaining of actions in MIA, it is conceived of in MCA as an interactive and reflexive practice that develops step-by-step in shifting micro-contexts.

In order to show the emergent and incremental character of the social interaction in orchestra rehearsals, this paper adopts an MCA-approach to transcription, which is sufficiently enriched by multimodal data. However, we adopt from MIA the central tenet of chained action types constituted by modal configurations and of differing intensities of the modes. By way of this mixed-methods approach, we hope to do justice to the interactional and multimodal complexities of orchestra rehearsals.

3 Orchestra rehearsals framed by multimodal (inter)action analysis

This section applies the basic notions of multimodal (inter)action analysis to orchestra rehearsals, thus seeking to prepare the ground for the sample analysis (cf. Table 1).

3.1 Interactional context

Following the general logic of multimodal (inter)action analysis (Norris 2004, 2016), orchestra rehearsals can be viewed on two different levels. First and more generally, they represent or instantiate a ‘nexus of practice’ (Scollon and Scollon 2004), which is “the point where multiple discursive and non-discursive practices come into contact and interact” (Norris and Maier 2014: 392). In this sense, orchestra rehearsals involve the historically shaped non-discursive practice of ‘translating’ sheet music into musical sound for performance and the multimodal communicative interaction (a discursive practice) of discussing, correcting and improving the orchestral play. A nexus of practice does not only include various types of discourse (cf. Scollon and Scollon 2004: 173), it also comprises the roles and identities of the actors and their interactional/discursive histories. The two most significant shaping factors in orchestra rehearsals as a nexus of practice are the ‘philosophy’ and approach of the conductor to the music and/or the rehearsing and the more or less established rules of practice and interaction for rehearsals in the given social set-up (e.g., conductor-in-chief vs. guest conductor).

Table 1: The conceptual framework of multimodal (inter)action analysis applied to orchestra rehearsals.

Interactional context		
TERM	GLOSS OF DEFINITION	APPLICATION TO ORCHESTRA REHEARSALS
<i>Nexus of practice</i>	non-discursive and discursive practices	music-making and music-commenting
<i>Site of engagement</i>	socio-physical setting for practice and analysis	orchestra rehearsal interaction
Action types		
TERM	GLOSS OF DEFINITION	APPLICATION TO ORCHESTRA REHEARSALS
<i>Chain of higher-level action</i>	sequence comprised of higher-level actions	instructing/instructional sequence
<i>Higher-level action</i>	basic action employing semiotic modes	e.g., locating, demonstrating etc.
<i>Lower-level action</i>	smallest pragmatic meaning unit of a mode	e.g., index finger upwards (gesture)
<i>Frozen action</i>	material object embodying previous actions	Score
Modal configuration		
TERM	GLOSS OF DEFINITION	APPLICATION TO ORCHESTRA REHEARSALS
<i>Modal configuration</i>	multimodal constitution of higher-level actions	sequencing and clustering of modes
<i>Modal complexity</i>	number of semiotic modes employed	from low (one) to high (four modes)
<i>Modal intensity</i>	relative importance of a semiotic mode	leading mode of singing/vocalizing

On a second, more specific level, orchestra rehearsals become a ‘site of engagement’, i.e., an analytical “window” onto “multiple converging practices” (Norris 2016: 123) and concrete actions. In this sense, orchestra rehearsals are primarily a practice of rehearsing music for the purpose of giving the score of a piece of music a unique sonic interpretation. Rehearsing generally involves repetition, corrective instruction and potentially a discussion of the envisaged interpretation itself – this is the broad actional scope of the site of engagement.

3.2 Action types

It is customary in multimodal (inter)action analysis to differentiate three types of action and to look at their inter-connections: lower-level, higher-level, and frozen actions (cf. Table 1). First, a lower-level action is taken to be the “smallest pragmatic meaning unit of a mode” (Norris 2019: 41), “in which a social actor draws upon a communicative mode such as gesture, posture, spoken language or layout, and constructs meaning” (Norris 2009: 82). For example, in an orchestra rehearsal, a conductor may raise his index finger, a minimal gestural unit that can signify a call for attention.

Second, many such lower-level actions are chained to produce higher-level actions. A higher-level action is constituted from “the coming together of a multitude of chains of lower-level actions” (Norris 2016: 125). For example, when, in orchestra rehearsals, conductors demonstrate the envisaged sonic quality of a piece of music to be replayed by the musicians, they perform a higher-level action. This, in turn, is composed by a number of lower-level actions, such as commenting, gesturing, directing the gaze and singing a passage. Higher-level actions may, on principle, utilize one semiotic resource or many of them in combination. The focus of the present study is on the types of higher-level actions in orchestra rehearsal and their modal constitution.

Third, multimodal interactions may involve what has been called ‘frozen action’, i.e., “chains of lower-level actions that are entailed in material objects within the site of engagement” (Pirini 2014: 80). In orchestra rehearsals, such an object that freezes previous action into an artifact and becomes pivotally involved in the interaction is the score of the piece of music to be rehearsed. Scores generally are notations of music as envisaged by a composer and contain a multitude of graphic marks that guide the music-making by providing concrete instructions for playing (e.g. dynamics, phrasing, bowing, special effects etc.). These notations shape both the playing and the ‘negotiation’ of certain sonic qualities of the music in the interaction between conductor and musicians. If we accept frozen action as “a useful methodological tool for the analysis of material objects” (Pirini 2014: 82) involved in interaction, we may want to integrate parts of the score into the multimodal transcript in order to show which specific frozen actions become relevant in the interaction.

3.3 Modal configuration

Contemporary multimodality research is primarily oriented towards mode linking and multimodal coherence (Stöckl 2019: 53–58, 64). In multimodal (inter)action analysis, the notion or “methodological tool” (Pirini 2014: 83) of ‘modal configuration’ or ‘modal density’ (Norris 2009: 84–88, 2016: 125; Pirini 2014: 83) accounts for this significance and covers “the importance of particular modes to an interaction, and how they interrelate”. Modal density as a super-ordinate concept comprises ‘modal complexity’ and ‘modal intensity’ (cf. Table 1). The first designates “the interrelationship of modes in a particular action” (Pirini 2014: 83), which can be glossed as the number of modes involved in any one higher-level action. Locating parts of the score to be played tends to have a low modal complexity as it can rely on speech alone, while demonstrating the sonic quality of the music shows a high modal complexity as it employs a number of modes such as vocalizing, gesture and body posture etc. ‘Modal intensity’, on the other hand, captures “how strongly one particular mode is engaged in during an interaction” (Pirini 2014: 83). We may ask about orchestra rehearsals, for instance, whether singing/vocalizing has a higher relative intensity than speech or gesture.

What emerges from establishing the types and number of modes involved in higher-level actions and their relative importance is an idea of how combinations of modes contribute to building higher-level actions (Norris 2009: 79). This has summarily been called ‘modal configuration’. It describes “the interplay of communicative modes as they are structured in relation to one another within a higher-level action that a social actor performs” (Norris 2009: 84).

Modal complexity and intensity have been linked to levels of awareness and degrees of attention (Norris 2019: 44, 243–251; Pirini 2014: 84–86) in as much as they “display the level of attention that a person pays in an (inter)action” (Norris 2019: 244). It would seem likely that in orchestra rehearsals, too, interactional moments that require special attention by the musicians are foregrounded or made more salient in perception by the conductor through high modal complexity (i.e., the use of many modes in parallel) and high modal intensity (i.e., the use of one central or leading mode). Consequently, musicians become more aware of certain higher-level actions than others and will likely pay more attention to them. In a pragmatic light, actions marked by high modal complexity or intensity have a stronger illocutionary force, which impacts on attention.

The approach in this paper elaborates on two central aspects of the theory of multimodal (inter)action analysis in particular: First, demarcating higher-level actions (Norris 2019: 189–194) is a crucial step to producing “bundled higher-level mediated actions” (Norris 2019: 193) and chains of higher-level actions. However, it backgrounds the ways in which the various higher-level actions are linked or sequenced into a higher-level action chain. We therefore propose that certain lower-level actions, for example different types of gazes (at the orchestra, at the score), serve to transition between the higher-level actions and function as pivots in a chain of higher-level actions. Second, the analytical tool of modal density (Norris 2019: 242) helps determine the configuration and relative importance of semiotic modes in higher-level actions and chains of higher-level actions. We take this and the observations on our data as the basis to distinguish between two basic types of modes: Default modes occur in all higher-level actions and thus foundationally carry the entire chain of actions (e.g., speech in rehearsals), whereas action-specific modes (e.g., vocalizing in rehearsals) occur more rarely and are restricted to one or few higher-level actions. In this way, default modes command less attention and awareness, while action-specific modes have a high salience and impact on attention more.

4 Methods

A preliminary analysis of various transcripts in the corpus (25-h of video-recording rehearsals in two French symphonic orchestras) suggests a delineation of a number of meaningful higher-level actions all constitutive of the chain of actions instructing how to play the music so as to create a model performance (cf. Table 2). This repertoire may not be complete but it does represent the core interactional work in orchestra rehearsals, showing that the main impetus for action constituted in whichever mode comes from the conductor.

Table 2: Repertoire of higher-level actions in orchestra rehearsals and their modal constitution – cf. transcript in Figure 1 below.

N°	HIGHER-LEVEL ACTION	GLOSS OF DEFINITION	MODES
1	<i>Interrupting</i>	halting the orchestra’s music-playing	<i>speech/gesture/gaze</i>
2	<i>Evaluating</i>	appraising/criticising the orchestra’s playing	<i>speech/gaze</i>
3	<i>Justifying</i>	announcing a critical/corrective commentary	<i>speech/gesture/gaze</i>
4	<i>Addressing</i>	identifying an orchestra section to be instructed	<i>speech/gaze/body movement</i>
5	<i>Locating</i>	identifying a part of the score to be played	<i>speech/gaze</i>
6	<i>Demonstrating</i>	illustrating the envisaged musical qualities	<i>speech/gesture/gaze/vocalizing</i>
7	<i>Clarifying</i>	addressing musician(s) and exchanging ideas	<i>speech/gaze/body movement</i>
8	<i>Describing</i>	characterizing the envisaged musical qualities	<i>speech/gesture/gaze</i>
9	<i>Signaling</i>	flagging a new phase of music-playing	<i>speech/gesture/gaze</i>

With a view to the corpus and regarding the exemplary transcript, four central sets of questions would seem worthwhile asking in a multimodal (inter)action analysis of orchestra rehearsals:

- (1) *Modal Complexity*: Which modes are used to constitute a given higher-level action? Which higher-level actions are modally more complex and which ones are less complex? What overall repertoire of modes emerges as typical of orchestra rehearsals?
- (2) *Modal Intensity*: In a given higher-level action, can a functionally central, leading or dominant modal resource be identified? What properties of a modal resource, i.e., qualities related to modal reach, explain its high intensity in the higher-level action?
- (3) *Default Mode*: Which modes appear to be a given in the higher-level actions of orchestra rehearsals? What qualities regarding their reach make them apparently indispensable? By contrast, which modes seem salient in the sense that they only occur in a specific higher-level action but not across the board?
- (4) *Transition Periods*: Do the transcripts allow the characterization of some modes as functioning to transition between higher-level actions? This would mean that certain modes are not part and parcel of a higher-level action but serve as pivots between higher-level actions, thus facilitating a smooth linking of the actions into a chain of higher-level actions.

In order to answer these questions, the transcript was analyzed applying the following steps in a methodological rationale (cf. Table 3). First, the interaction was segmented into potentially minimal units, i.e., higher-level-actions, applying the repertoire in Table 2. This forms the prerequisite for statements about the structure of the concrete interaction that emerges from different higher-level actions. Second, each higher-level action was interrogated for its modal complexity, that is for the types and number of semiotic modes constitutive of the action. This analytical step helps form ideas about similar or varying degrees of modal complexity in the semiotic repertoire of the individual actions. It also forms the basis for observations about the functional impact of each modal resource for the performance of the individual action.

Third, we were interested in whether there is something like a central or dominant semiotic resource in any of the higher-level actions. Such modes with a higher modal intensity would be leading the constitution of the higher-level action and have more relative impact on it. This methodological step entails asking about and comparing the reach of the various modes in the individual actions. Fourth, by comparing the repertoire of semiotic modes used across all higher-level actions, ideas can be formed about modes that feature throughout the action chain of instructing and about modes that appear unique and typical of one higher-level action only.

Finally, the analysis focused on how the higher-level actions are sequenced into the instructional chain of higher-level actions. Here, we wanted to find out whether perhaps some modes take over a specific function of transitioning between higher-level actions, instead of constructing the higher-level action itself. This is the moment in the analysis where special attention is devoted to the function of frozen action in the form of the score. It appears that transitions are crucially linked to the score as it has a central function for everyone involved in the logic of the interaction.

We will now present and interpret the transcript of the rehearsal fragment. As a result, we can ascertain the sequence of the higher-level actions as they emerge from the ongoing interaction. The transcript presentation will then be used to discuss the results obtained from an application of the five steps in our methodological rationale.

Table 3: Five steps in the methodological rationale of the study – as applied to the transcript.

STEP N°	OBJECTIVE	RESULTS
1	segment the interaction into higher-level actions	sequence/structure
2	fix types and number of semiotic modes in the actions	modal complexity of actions
3	ascertain central/dominant modes	modal intensity
4	compare actions regarding modes	default versus action-specific modes
5	check transitioning between higher-level actions	function of frozen action

5 Transcript analysis and interpretation

5.1 Data and multimodal transcription

The short extract chosen here (cf. Figure 1) is 22 s from a rehearsal of the *Orchestre de Paris* with the Italian guest conductor Gianandrea Nosedà. In the snippet, the orchestra is rehearsing *Messa da Requiem* by Verdi. The languages used are French, English and Italian.

```
((musicians play and conductor conducts))

01  C  ♪&okay!♯ ((claps)) c'est bon! *(1.0)*
                                     it's good
      c  ♯stops conducting♯
          &looks at the score --->(03)
                                     *extends right arm and right index
                                     forwards to the right*

      mm ♪((are playing)) --->(03)

02  C  (0.5) ka- ~eh:: (0.4) eh::~~
      c  ~holds arms chest-high~

03  C  (.) *why i'm asking* ~you,& ♯why i'm asking you~♯ portato lungo.♪
      c  *turns hands forwards in a circular motion*
          ~moves fingers ----->~
          ---->& ♯looks to the left♯

      mm                                     ----->♪

04  C  &°h:: ~ (0.5)& *parce que (.) nous pouvons~
                                     because we can
          &looks at the score&
          *looks to the left ---->(07)
          ~.....turns hands forwards in a circular motion~

05  C  (0.5) &#<<1> tam pam pam pam [pam pam& pa-]>
          &moves upper body and arms back and forth&

      fig      #fig.1a


06  CM                                     [xxx      ]

07  C  &si ah je sais&* ~eh?
      yes i know
      &nods ----->&
          ---->* ~looks at the score ---->

08  °h et &#le~ *le: (0.9)
      and the
          ---->~ *looks to the orchestra ---->
          &extends indexes upwards ---->
```

Figure 1: Transcript Orchestre de Paris, 1_09022016, 0008, 03:41–04:03 (a) Movement of upper body and arms. (b) Indexes extended upwards. (c) Imitating the bow movement.

fig #fig.1b



```

09  C   le: (.) the IDEA is&  just to MAKE the~* ~phrasing
      --->&  turns hands forwards in a circular motion~
      --->* ~looks at the score --->

10      &#<<1> sol sol sol sol sol sol&~
      imitates bow movement with hands&
      ----->~

```

fig #fig.1c



```

11  C   &*<<h> la la> <<gets lower> sol sol fa fa mi>&*
      looks to the left&
      *moves upper body and arms back and forth --->*

12      &*eh? (0.7) <<f> nous le farons!> nous le farons.
      we will manage it
      looks at the score ----->
      *leaves through the score ----->

```

Figure 1: Continued.

For the transcription, we have adopted a conversation analytic approach which is based on GAT II-conventions for talk (cf. Selting et al. 2009) and on Mondada's (2014) conventions for multimodality. We are working with the temporal location and description of embodied actions as well as with screenshots integrated in the transcription. This is important for the identification of temporally and sequentially organized details of actions.

5.2 Extract: *parce que nous pouvons tam pam pam pam*

In this extract the conductor first interrupts the music using verbal and body cues. Then he directs the musicians to play a certain passage in the score with a specific articulation (*portato lungo*) and asks them to concentrate on the linking between the notes (*phrasing*). In doing so, he applies a variety of different modes.

This sequence shows a succession of higher-level actions that constitute the instructional chain of higher-level actions. The passage begins with an interruption of the music by the conductor in L01 and L02, followed by an explanation for this interruption and a motivation for the upcoming instruction (L03). In a third step, the conductor verbally introduces a directive (L04) which he completes vocally (L05). While the conductor sings a

passage from the score (L05), the concertmaster intervenes with an unintelligible turn (L06); in L07 the conductor reacts in a confirmational way to this turn. After this short kind of detour sequence with the concertmaster, the conductor continues his instruction verbally (L08 and L09) before he again switches to singing in L10 and L11. In L12, he concludes the sequence with a tag question and an inclusive-we request for the musicians to play the passage commented on.

This short sequence contains different higher-level actions:

- 1) interrupting/evaluating (L01–02),
- 2) accounting/justifying (L03),
- 3) introducing verbally an instruction (L04),
- 4) vocal directing/demonstrating (L05),
- 5) clarifying (L06–07),
- 6) describing (L08–09),
- 7) vocal directing/demonstrating (L10–11),
- 8) and signaling (L12).

In addition, it is possible to identify two further higher-level actions, which are realized in an implicit way. In L03, as the conductor uses the technical expression *portato lungo*, he refers to a passage in the score where this technique of playing the music occurs (marked in green, cf. Figure 2).

In this way, the conductor locates and identifies a certain passage in the score that is going to be discussed and becomes the object of instruction (L04–12). At the same time, he implicitly addresses certain parts of the orchestra, that is, all the musicians to whom the score allocates a *portato*, namely first and second violins, violas and cellos (marked in yellow).

In this short extract, the conductor applies different semiotic modes for interrupting, addressing, locating and explaining. For interrupting, he uses verbal, body and gaze cues (L01), for addressing he makes use of gesture and gaze (L01 and L03), for locating he employs a technical expression (L03), and for justifying he asks a question (verbally), moves his hands (gesturally) and looks to the left (gaze). Interrupting and justifying can be recognized as the most complex actions. For these two activities, the conductor makes use of three different semiotic modes, while for addressing and locating he employs one mode at a time. This could be connected to the fact that the principal objective of the conductor in this sequence is to arrest the playing of the orchestra. In fact, the musicians take time to stop the music until the end of the conductor's explanation in L03. Thus, both the conductor's attempt to interrupt as a *standout* action and the justification as an anticipatory activity to the following instructions are realized with more effort and through a complex interplay of different modes as compared with addressing and locating. The higher-level action of interrupting is here described as *standing out* because of its high illocutionary force in the structure of the rehearsal. *Standout* actions have a rich semiotic content, they are constructed in an interplay of different modes. These characteristics enable a more effective expression of their function, i.e. their illocutionary force.

Figure 2: Extract from the score I.

In L04–L12, the conductor gives an instruction connected to a passage identified in the score before (*portato*). In L04 he answers his question from L03 verbally (*parce que nous pouvons*, ‘because we can’) and completes the turn by a vocally demonstrated action in L05 (*tam pam pam pam pam pam pa*). While doing so, he turns his hands forwards in a circular motion (as in L03), he looks first at the score and then again to the left, in the direction of the violins. This orientation of the gaze continues until L07 when the conductor terminates the first part of his directives with a tag question (*eh?*). The content proper of the instruction is expressed in L05 through a vocal imitation of the *portato* notes in the violins, violas and cellos. While singing, the conductor moves his upper body and his arms back and forth (cf. Figure 1a).

The combination of a verbal introduction and an embodied completion of a turn can be described as a “syntactic-bodily gestalt” (Keevallik 2015: 309). The conductor, by imitating the performance of the musicians (through singing or gesturing), is able to judge and modify the music played by the orchestra and to give the musicians quite a precise idea of how a certain passage in the score should be read and performed. In this combination of different semiotic modes (e.g., talk, singing, gesture), talk according to Keevallik (2015: 310) serves as a “keying device” or rather as an introductory/framing advice for the following embodied demonstration. In this perspective, in such *syntactic-bodily gestalts*, the leading modes are singing and gesture as they instruct the musicians on how to play something in the next part of the performance.

In L06, the concertmaster interrupts the sung and gestural demonstration of the conductor with a verbal (unintelligible) gambit, to which the conductor replies in L07 with an affirmative verbal turn (*si ah je sais*, ‘yes I know’) and by nodding. This short side-sequence (L06–07) is inserted in the ongoing activity of instructing. It shows that in the hierarchical setting of the orchestra rehearsals, some actors like the concertmaster have the right to take the floor and to enter into a one-face-to-one-face communication mode with the conductor (but still with the other musicians as an audience).

After this question-answer sequence with the concertmaster, the conductor again focuses his gaze on the score (L07). In L08, he continues his instruction (introduced by the conjunction *et*, ‘and’), shifts his gaze to the orchestra and extends his index fingers upwards (cf. Figure 1b). With this gesture, the conductor illustrates the speech act of warning or paying attention. In this way, he makes assertive claims that reinforce the ones realized by speech. In fact, the conductor verbally explains his view (*the idea is just to make the phrasing*, L09), and it is only by using the coverbal gesture that the verbal turn acquires instructional meaning.

This index position (cf. Figure 1b) lasts until the conductor in L09 again turns his hands forwards in a circular motion (similarly to L03 and L04). He also shifts his gaze away from the orchestra towards the score and begins to sing in L10 (*sol sol sol sol sol sol*). For this vocal passage, the conductor uses Italian tone syllables (*sol*), in contrast to L05, where he employs freely selected syllables (*tam* and *pam*); this way of singing even continues in L11 (*la la sol sol fa fa mi*). It becomes clear that by using these specific syllables (*sol*, *la*, *fa*, *mi*), the conductor refers to the second and third bar of the passage located before in the score (marked in blue, cf. Figure 3).

Figure 3: Extract from the score II.

L10–11 do not only show a different way of singing, but also a contrasting way of instructing. In L03–05, the conductor first gives a motivation for the following instruction and then produces a *syntactic-bodily gestalt*. In L09–11, he first gives a verbal directive followed by an embodied instruction that develops the same content (the musical ‘phrasing’). The embodiment in L10 is made by gesturing and singing: the conductor sings a violin passage from the score and moves his hands imitating the bow stroke (cf. Figure 1c). Here the conductor uses iconic gestures such as the bow stroke and sung quotations (*sol sol sol sol sol sol*) to iconically demonstrate how to play. In other words, the conductor uses two different semiotic modes (singing and gesturing) in order to point out (or ‘highlight’, cf. Goodwin 1994) those elements that are in need of correcting and are to be heeded by the musicians when playing the next part of the music.

In L09–11 therefore, the conductor conveys the same content twice but only in different semiotic modes: 1) verbally and gesturally (L09), 2) by singing, imitating, gesturing and gazing (L10–11). After instructing the musicians on the three bars, the conductor concludes the discussion of this short passage with a tag question (*eh?*) and a projecting expression (*nous le faisons*, ‘we will manage it’) in L12. Simultaneously he looks at the score – and no longer at the orchestra or to the left, in the direction of the violins – and leafs through the pages. The shift in the conductor’s gaze and attention supports the conclusive character of his turn in L12.

6 Discussion of results

6.1 Modal complexity

The sequence of higher-level actions established for the transcript suggests a rich interactional structure. With six higher-level actions comprising three modes each and one higher-level action comprising 4 modes (cf. Table 2), the overall modal complexity appears to be high. This corroborates concrete evidence of Norris’ dictum that “no mode is ever used or developed in isolation” (Norris 2013: 163). Emerging from the medial situation (cf. Section 1), two types of semiotic modes may be distinguished: auditive (including speech and vocalizing) and visual (including gesture, gaze and body posture/movement) (cf. Table 4).

First, the combination of speech, gesture and gaze seems frequent and generic. This may be interpreted in the classic way of speech necessitating paraverbal semiotic modes. However, it may, in this special context, also be taken as evidence of the central role of conducting, which is essential to orchestral music-playing and might be seen to extend into the instructional interaction. Second, vocalizing adds to only one higher-level action, namely *demonstrating*, where it appears to take on a central role and seamlessly cooperates with speech, gesture and gaze.

It is our view here that each semiotic resource is involved in multimodal interaction for a reason, that is, it fulfills a function owing to and determined by its communicative strengths – a notion Kress (2010: 83) dubbed ‘modal reach’. Speech allows for abstraction and flexible reference, facilitates negotiating a virtually limitless range of concepts and can perform all kinds of speech acts. This may explain its central role in rehearsal interaction. However, when the sound qualities of music need to be illustrated, singing and rhythmic vocalization seem to be better suited to the job as they can directly and effectively imitate the music. Gesture has its communicative strengths in pointing to sections of the orchestra and in supporting the imitation and

Table 4: Modes and their modal reaches.

MODAL RESOURCE	TYPE OF MODE	MODAL REACH (IN REHEARSALS)
<i>Speech</i>	auditive	flexible reference to and speech acts about music
<i>Vocalizing</i>		direct (iconic) imitation of musical qualities
<i>Gesture</i>	visual	pointing to sections and illustrating tune/phrasing
<i>Gaze</i>		address and orientation towards score
<i>Body posture/movement</i>		address and orientation towards orchestra sections

illustration of tunes and rhythmic phrasing. Gaze manages interpersonal rapport and address, but it also serves to indicate an involvement of the conductor with the score. Such an involvement is fundamental for his orientation as it helps him organize the instruction and read the frozen action for what it entails for the musical practice. Body posture/movement, finally, are visual semiotic modes that, similar to gaze, can indicate and underscore an action of addressing individual musicians or sections of the orchestra. Consequently, they occur in the higher-level actions of clarifying and addressing. It would seem plausible to assume that the reaches of the semiotic modes employed in orchestra rehearsals may be used to model potential modal intensities (cf. Section 6.2 below).

6.2 Modal intensity

Determining “which modes are absolutely necessary so that the social actor can perform the action(s)” (Norris 2009: 83) is not a straightforward task but fraught with difficulty. The concept of ‘modal intensity’ echoes the notion of ‘relative status’ as discussed in multimodality research generally (Stöckl 2020: 190–195); modes are said to have equal or unequal status. When one mode is central, leading or dominant in a combination of modes, i.e., when it has a higher modal intensity, we assume they are unequal. The reverse case would be a combination of modes where each has the same intensity so that either modes are in a complementary relation or function independently of one another.

The transcript analysis suggests that in orchestra rehearsals indeed some modes seem to take on a central role; these are *speech* and *vocalizing*. Regarding speech and gesture, the observations tell us that these are pervasive modes, which occur in a large number of higher-level actions, often in conjunction with each other. As for vocalizing, this modal resource is unique to the higher-level action of *demonstrating*, but there it clearly performs a central function. We take the general view here that mode-centricity or modal intensity derives from modal reaches and can be explained on the basis of what a given modal resource contributes to the action.

While it may not be temporally dominant, *speech* features in all higher-level actions and is central because it ties together most modes in a complementary fashion. The high intensity of speech rests on its potential to express speech acts clearly, a function no other mode can perform as effectively. In the orchestra rehearsal these would be: praising/criticizing, justifying, confirming, explaining, identifying, locating and initiating. Another vital semiotic work that is efficiently performed by speech due to its precise and flexible reference is designating parts of the score, specific instructional notes, and sections of the orchestra.

Vocalizing, on the other hand, is very well suited to demonstrating or clarifying certain structural, rhythmic or sonic qualities of the music. This illustrating function relies primarily on the iconic qualities of singing/humming, which can immediately imitate musical qualities without having to ‘translate’ meaning from one mode (music) to another. In the inherently and strongly multilingual practice that contemporary orchestras present, sidestepping speech may also be seen to solve a problem of cross-cultural, interlingual communication.

Gesture, finally, does not qualify as a central modal resource, even though it features in five of the nine higher-level actions. Transcript evidence suggests that gesture is rather in complementary relation to speech. In this modal combination, it serves to support the pointing to certain orchestra sections, the halting/resuming of the playing or the demonstration and description of musical qualities.

6.3 Action-specific modes

It would seem plausible that modes that pervade the entire chain of higher-level instructional actions constitute some kind of default mode, whereas modes that only occur rarely are higher-level-action-specific and gain in salience. In this view, speech, gesture and gaze qualify as default modes as they form a modal combination in four of the nine higher-level actions, namely interrupting, justifying, describing, and signaling.

By contrast, vocalizing and body posture/movement only occur in few of the higher-level actions and would, therefore, appear to be action-specific modes for demonstrating, addressing and clarifying.

While we have seen that gesture fulfills specific functions in orchestra rehearsals (cf. Section 6.2 above), the modal default combination of *speech – gesture – gaze* can still safely be interpreted as medium-specific and generic. In face-to-face communication, speech necessitates the paraverbal modes of gesture and gaze, which may, however, undergo a contextual specification of their functions. Vocalizing, on the other hand, seems highly typical of orchestra rehearsals as it fulfills context-specific functions in only one, albeit central higher-level action (i.e., *demonstrating*); it can, therefore, be seen to be salient. The same holds for *body posture/movement*, which is utilized in only two of the nine higher-level actions, namely in *clarifying* and *addressing*. In both, body movement underscores the addressing of the orchestra or sections of it by the conductor, so that a complementary relation between speech/gaze and body movement is achieved.

While *gaze* seems to come part and parcel of speech, it must also be reconsidered in the light of its function as a modal resource for the conductor to orient towards the score/the frozen action (cf. Section 6.4 below). In summary, then, default modes fulfill various recurrent semiotic functions in a chain of higher-level actions, whereas action-specific modes are tied to specific higher-level actions as they perform specialized and salient semiotic work.

6.4 Frozen action and transitioning between higher-level actions

When watching the orchestra rehearsals in the corpus, gaze appeared to be omnipresent, an observation born out in the sample-transcript and its modal configuration of the actions. Rather than simply be part of the modal combination of all specific higher-level actions, gaze had better be interpreted as fulfilling a function with regard to the smooth linking or sequencing of the actions into the instructional chain of higher-level actions. Such a view endorses gaze as a modal resource that acts as a pivot for the conductor or a transitioning device to switch between actions and organize the structure of the rehearsal. The gaze of the conductor can either be directed at the score or at sections of the orchestra. While the latter seems tied up with the function of addressing a selection of musicians, the former focuses attention on the frozen action.

Scores represent the frozen action in orchestra rehearsals as they contain the substance of the music to be played as well as instructions (in the form of additional notation) on how the music is to be performed. For both, musicians and conductor, the score is the most central material object of a rehearsal as it allows all actors to read and follow the music simultaneously. By frequently looking at the score, the conductor signals the importance of the frozen action and utilizes it to glean information about specific parts of the music and its interpretation. Conductor's and musicians' actions converge on the score as successful instruction crucially results in note-taking in the score and the sheet music in order to save in writing the ways in which the music is to be played. In conclusion, instances of gaze become an independent transitioning device. They signal moves from one higher-level action to the next, and by connecting with the score and the orchestra, gaze helps realize the ideational and the interpersonal moments in orchestra rehearsals.

7 Conclusions

Applying multimodal (inter)action analysis to a sample-transcript, the present contribution has provided some first insights into the actional structure and modal configuration of orchestra rehearsals (cf. Figure 4). Over and above the repertoire of higher-level actions in musical instruction, special attention was devoted to the modal complexity of the actions and the varying intensity of the modes at work. We extended the framework of analysis in two ways: by introducing a distinction between default and action-specific modes and by arguing that modes may act as pivots between actions to facilitate the smooth linking of actions into a chain. Modal complexity overall was found to be high, with speech, vocalizing, gesture, gaze and body posture/movement constituting nine distinct higher-level actions. Owing to their significant functions, speech and vocalizing were

found to exhibit a high modal intensity. While speech, gaze and gesture appeared to be all-pervasive default modes, vocalizing and body posture proved to be action-specific in different ways. Finally, gaze turned out to function as a pivot between higher-level actions, facilitating a smooth transition between actions by providing the conductor with visual orientation towards the score/frozen action and towards the musicians.

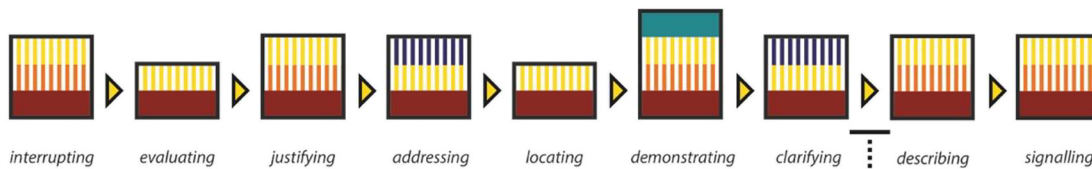
Given that more transcript material was studied applying the rationale of this paper, the multimodal interaction of orchestra rehearsals could be refined in two ways: First, our preliminary observations about the interactional structure and modal configuration of the rehearsals would be ascertained, enhanced and modified. Second, once the analytical procedure was verified, one could finally turn the attention to the variability of multimodal interaction in orchestra rehearsals. A number of obvious social factors relating to the nexus of practice would commend themselves for the study of variation: multilingual status, i.e., the number of languages spoken to what extent; orchestra-conductor relation, i.e., long-standing practices versus newly established or shifting practices; and professional culture, i.e., instructional/interactional practices that have become conventionalized for a given orchestra or a (national) orchestra culture.

CHAIN OF HIGHER-LEVEL ACTIONS

Visualisation: Jana Pflaeging & Hartmut Stöckl

..... playing instructing feedback/discussion playing▶

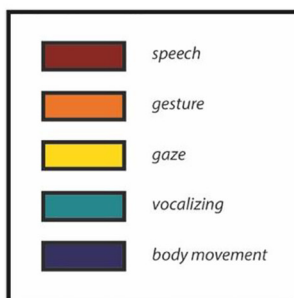
HIGHER-LEVEL ACTIONS



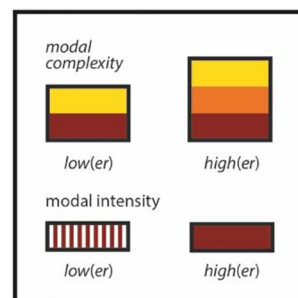
FROZEN ACTION



SEMIOTIC MODES



MODAL CONFIGURATIONS



TRANSITIONS

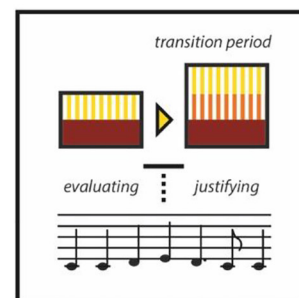


Figure 4: Modeling multimodal interaction in orchestra rehearsal – levels and relations.

Acknowledgments: We would like to thank the Orchestre de Paris and its Italian guest conductor Gianandrea Noseda for their permission to make the recordings of the rehearsals and to utilize these in the present paper. This includes the permission to reproduce the three stills and the transcript of the video.

References

- Adenot, P. (2015). The orchestra conductor. From the authority figure to negotiated order in a vocational profession. *Transposition* 5: 2–14.
- Bateman, J., Wildfeuer, J., and Hiippala, T. (2017). *Multimodality: foundations, research and analysis: a problem-oriented approach*. de Gruyter, Berlin/Boston.
- Biasutti, M. (2012). Orchestra rehearsal strategies: conductor and performer views. *Music. Sci.* 17: 57–71.
- Bräm, T. and Boyes Bräm, P. (2004). A pilot study of the expressive gestures used by classical orchestra conductors. In: Emmorey, K. and Lane, H. (Eds.), *The signs of language revisited: an anthology in honor of Ursula Bellugi and Edward Klima*. Lawrence Erlbaum, Mahwah, pp. 143–168.
- Clark, H.H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123: 324–347.
- Clark, H.H. and Gerrig, R.J. (1990). Quotations as demonstrations. *Language* 66: 764–805.
- Deppermann, A. (2013). Multimodal interaction from a conversation analytic perspective. *J. Pragmat.* 46: 1–7.
- Deppermann, A. (2018). Sprache in der multimodalen Interaktion. In: Deppermann, A. and Reineke, S. (Eds.), *Sprache im kommunikativen, interaktiven und kulturellen Kontext*. de Gruyter, Berlin/Boston, pp. 51–85.
- Deppermann, A., Feilke, H., and Linke, A. (2016). Sprachliche und kommunikative Praktiken: eine Annäherung aus linguistischer Sicht. In: Deppermann, A. and Reineke, S. (Eds.), *Sprachliche und kommunikative Praktiken*. de Gruyter, Berlin, pp. 1–23.
- Faulkner, R.R. (1973). Some features of communication and authority in an artistic organization. *Sociol. Q.* 14: 147–157.
- Goffman, E. (1974). *Frame analysis: an essay on the organization of experience*. Northeastern UP, Boston.
- Goodwin, C. (1994). Professional vision. *Am. Anthropol.* 96: 606–633.
- Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *J. Pragmat.* 46: 8–23.
- Haviland, J.B. (2007). Master speakers, master gestures: a string quartet master class. In: Duncan, S.D., Levy, E.T., and Cassell, J. (Eds.), *Gesture and the dynamic dimension of language: essays in honor of David McNeill*. Benjamins, Amsterdam, pp. 147–172.
- Ivaldi, A. (2016). Students' and teachers' orientation to learning and performing in music conservatoire lesson interactions. *Psychol. Music* 44: 202–218.
- Keevallik, L. (2010). Bodily quoting in dance correction. *Res. Lang. Soc. Interact.* 43: 401–426.
- Keevallik, L. (2015). Coordinating the temporalities of talk and dance. In: Deppermann, A. and Günthner, S. (Eds.), *Temporality in interaction*. Benjamins, Amsterdam et al., pp. 309–336.
- Kendon, A. (2004). *Gestures: visible action as utterance*. CUP, Cambridge.
- Kress, G. (2010). *Multimodality: a social semiotic approach to contemporary communication*. Routledge, London.
- Mondada, L. (2013). Multimodal interaction. In: Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., and Tessendorf, S. (Eds.), *Body – language – communication: an international handbook on multimodality in human interaction*. de Gruyter, Berlin, pp. 577–589.
- Mondada, L. (2014). Conventions for multimodal transcription, Available at: https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf (Accessed 16 December 2020).
- Mondada, L. (2016). Challenges of multimodality: language and the body in social interaction. *J. SocioLinguistics* 20: 2–32.
- Nishizaka, A. (2006). What to learn: the embodied structure of the environment. *Res. Lang. Soc. Interact.* 39: 119–154.
- Norris, S. (2004). *Analyzing multimodal interaction. A methodological framework*. Routledge, New York/London.
- Norris, S. (2009). Modal density and modal configurations: multimodal actions. In: Jewitt, C. (Ed.), *The routledge handbook of multimodal analysis*. Routledge, London, pp. 78–90.
- Norris, S. (2011). *Identity in (inter)action. Introducing multimodal (inter)action analysis*. de Gruyter, Berlin/Boston.
- Norris, S. (2013). What is a mode? Smell, olfactory perception, and the notion of mode in multimodal mediated theory. *Multimodal Commun.* 2: 155–169.
- Norris, S. (2016). Multimodal interaction: language and modal configurations. In: Klug, N.-M. and Stöckl, H. (Eds.), *Handbuch Sprache im multimodalen Kontext*. de Gruyter, Berlin/Boston, pp. 121–142.
- Norris, S. (2019). *Systematically working with multimodal data: research methods in multimodal discourse analysis*. Wiley Blackwell, Hoboken, NJ.
- Norris, S. and Maier, C.D. (Ed.) (2014). *Interactions, images and texts: a reader in multimodality*. de Gruyter, Boston/Berlin.
- Pirini, J. (2014). Introduction to multimodal (inter)action analysis. In: Norris, S. and Maier, C.D. (Eds.), *Interactions, images, and text: a reader in multimodality*. de Gruyter, Berlin/Boston, pp. 77–91.

- Price, H.E. and Byo, J.L. (2002). Rehearsing and conducting. In: Parncutt, R. and McPherson, G. (Eds.), *The science and psychology of music performance: creative strategies for teaching and learning*. Oxford University Press, Oxford, pp. 335–351.
- Scollon, R. and Scollon, S.W. (2004). *Nexus analysis: discourses and the emergent internet*. Routledge, London/New York.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günthner, S., et al. (2009). Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforschung – Online-Z. zur verbalen Interaktion* 10: 353–402.
- Stevanovic, M. and Kuusisto, A. (2018). Teacher directives in children’s musical instrument instruction: activity context, student cooperation, and institutional priority. *Scand. J. Educ. Res.* 63: 1022–1040.
- Stöckl, H. (2019). Linguistic multimodality – multimodal linguistics: a state-of-the-art sketch. In: Wildfeuer, J., et al. (Eds.), *Multimodality: disciplinary thoughts and the challenge of diversity*. de Gruyter, Berlin/Boston, pp. 41–68.
- Stöckl, H. (2020). Multimodality and mediality in an image-centric semiosphere: a rationale. In: Thurlow, C., Dürscheid, C., and Diémos, F. (Eds.), *Visualizing digital discourse: interactional, institutional and ideological perspectives*. de Gruyter, Berlin/Boston, pp. 189–202.
- Szcepek Reed, B., Reed, D., and Haddon, E. (2013). NOW or NOT NOW: coordinating restarts in the pursuit of learnables in vocal master classes. *Res. Lang. Soc. Interact.* 46: 22–46.
- Tolins, J. (2013). Assessment and direction through nonlexical vocalizations in music instruction. *Res. Lang. Soc. Interact.* 46: 47–64.
- Veronesi, D. (2014). Correction sequences and semiotic resources in ensemble music workshops: the case of conduction®. *Soc. Semiotic.* 24: 468–494.
- Weeks, P. (1996). A rehearsal of a Beethoven passage: an analysis of correction talk. *Res. Lang. Soc. Interact.* 29: 247–290.