# Papers of the Linguistic Society of Belgium

## Contrast Pairs in Orchestra Instruction as a Window into the Multimodal Expression of Stance

Katharina Meissl, Paul Sambre, Kurt Feyaerts

# Contrast Pairs in Orchestra Instruction as a Window into the Multimodal Expression of Stance

Katharina Meissl, Paul Sambre, Kurt Feyaerts

*KU Leuven*

On the background of the increased interest in the phenomenon of stance-taking (i.a. Englebretson 2007, Debras 2015) this contribution explores the notion of *contrast pairs* (Weeks 1996, Messner 2020) as it is instantiated in orchestra instruction during rehearsal, an inherently stance-laden activity given its focus on evaluating and improving the musicians' collective performance. Contrast pairs are sequences of closely juxtaposed depictions (Clark 2016, or "illustrative expressions" in Weeks 1996), which serve the instructional comparison of undesired and desired (orchestral) performances. In this contribution, we report on an exploratory analysis of a corpus of approximately ten hours of video recordings of orchestra rehearsals (Schrooten & Feyaerts 2020). Through a close reading of two examples, we shed light on the sequential and multimodal complexity of contrast pairs and identify steps for further analysis. We argue that, due to their evaluative and contrastive nature, contrast pairs offer valuable grounds to contribute to the study of multimodal stance-taking in interaction.

## 1. Introduction

Recently, there has been an increasing interest in the study of stance, which can be defined as "attitude which, for some time, is expressed and sustained interactively in communication, in a unimodal or multimodal manner" (Chindamo *et al.* 2012:618). According to Du Bois (2007:163), stance-taking involves three simultaneously occurring processes: evaluating objects, positioning subjects and (dis)aligning with other subjects. A range of studies has investigated different communicative means by which stance-taking is achieved:

lexico-grammatical stance markers (Kärkkäinen 2007, Heritage 2012, Bohmann & Ahlers 2021), prosody and manual gesture (Borràs-Comes *et al.* 2019), head movement (Stivers 2008) shrugs (Debras 2017, Jehoul *et al.* 2017), eye rolls (Clift 2021) as well as combinations of several different markers (Debras 2015, de Vries *et al.* 2021) also involving body repositioning and gaze behaviour.

What emerges from this literature is that humans tend to constantly evaluate objects, states of affairs, people (and their stances) more or less explicitly and through different semiotic modes: "[…] there is never a time out from the social action of taking stances and adopting positions (Du Bois & Kärkkäinen 2012:438). Due to this omnipresence, stance can be used as a label for a broad range of phenomena already otherwise labelled according to their form, such as shoulder shrugs or modal verbs. Yet, only few studies have looked at how multiple semiotic modes work in concert to express stance.

Therefore, in order to make more informed statements about the forms that stance can take in interaction, especially when marked non-lexically, it is necessary to examine comparable instances thereof as a basis for further theoretical and methodological development. We argue that "contrast pairs" (Weeks 1996), which contain closely juxtaposed multimodal depictions (Clark 2016) that serve the instructional comparison of undesired and desired (orchestral) performances, present themselves as such distinctive and comparable expressions of stance in similar discursive contexts. They involve directly contrasted positive and negative evaluations realised through speech and complex multimodal conduct. By investigating this practice, we aim to add a piece to the puzzle of multimodal stance, also considering the added dimension of instruction and interaction on music as a semiotic mode.

The article is structured as follows: Section 2 deals with orchestra rehearsals as an instructional setting and the concept of contrast pairs. In section 3, the research aims, video corpus and annotation process are described and later exemplified with a close reading of two cases in section 4. Based on the discussion of the sequential and multimodal complexity of contrast pairs, further analytical steps are identified in section 5.


## 2.   Contrast pairs in orchestra rehearsals

Put simply, the goal of orchestra rehearsals is to improve the collective musical performance of an ensemble working towards a concert. This process aims at achieving maximal alignment between the conductor's intention and the musicians' performance of the written score of the respective musical piece (Stoeckl & Messner 2021:2). While the score serves as an important frame of

reference uniting both performative perspectives, the notation as such does not include all aspects relevant for the interpretation, such as dynamics (cf. Meissl *et al.* 2022). Thus, the meaning of the score emerges in the course of interaction during rehearsal, with the interpretation the conductor conveys to the orchestra through instructions (Weeks 1996:248).

To bridge the knowledge gap that opens up between conductor and musicians regarding the intended interpretation, conductors try to ensure understanding and intersubjectivity through their behaviour. As described in studies on a variety of instructional settings such as driving lessons (Deppermann 2018), dance lessons (Keevallik 2010), musical master classes (Reed & Szczepek Reed 2014) and orchestra rehearsals (Messner 2020), instructors tap into a multitude of semiotic resources. These include speech, non-lexical vocalisations, gesture, facial expressions, body posture and movement as well as gaze behaviour. A combination of different visual and auditory semiotic modes allows for the expression of complex, layered meanings and stances pertaining to abstract concepts like sound, atmospheres, emotions, or movement made relevant during music-making.

Conductors use different practices of depiction to make themselves clear. Ehmer (2013) summarises multiple verbal strategies such as giving examples or drawing on verbal imagery, for instance, describing a certain sound as a haptic sensation or an emotion. As conductors often refrain from using speech, offering embodied displays of musical passages with vocalisations and movement instead, we also draw on the concept of depiction, coined and defined by Clark (2016:325) as "physical scenes that people stage for others to use in imagining the scenes depicted". In the context of orchestra rehearsals, depictions are often included in instruction sequences mostly referring to the way in which music should (not) be played.

In the asymmetric and hierarchical setting of orchestra rehearsals, conductors are at liberty to give instructions at any given point during the rehearsal (Stoeckl & Messner 2021:2). These instructions occur in roughly two alternating types of sequence: conducting sequences, where the orchestra plays, and instructional sequences, where play is interrupted. Within these higher-level sequences, whose alternation can be compared to Initiation-Response-Evaluation sequences in classroom interaction (Sinclair & Brazil 1982), we discern several micro-level activities such as providing information, storytelling, demonstrating and depicting. In this paper, we focus on instructional instances located between sequences of musical performance, where conductors pause and discuss the musicians' performance in reference to specific points for improvement. We will refer to these as "learnables" (Reed & Szczepek Reed 2014, Zemel & Koschmann 2014), similar to what Schegloff et al. (1977), in their conversation analytic work on repair, categorise as "correctables", a term also used in studies

on performing arts instruction (Ivaldi *et al.* 2021:2, 7–8). Such learnables can be aligned both retro- and prospectively, either by critically commenting on a previous performance or by prescribing certain adjustments, which inherently overlaps with stance-taking acts due to their evaluative nature and goal of alignment. Multimodal depictions of musical passages and performance often form a crucial part of such instructions. By offering assessments and instructions, conductors assume a position in relation to the musicians and their performance (Du Bois 2007), making clear how the piece in question should or should not be played. Sometimes these formats of instruction are used in direct contrast with each other. Weeks coined such instances as "contrast pairs*"*:

> "[…] pairs of IEs [illustrative expressions] - one embodying the **faulted** performed version of a given musical passage, the other exemplifying the conductor's **prescribed** version. They follow closely on one another, though not necessarily immediately contiguously; there often is a short verbal introduction to the second one" (Weeks 1996: 269, highlighted by authors).

Such "illustrative expressions" are, in contrast to what he calls verbal expressions, "sung, chanted, or counted or […] other embodiments of the music" (Weeks 1996: 254). This type of expression corresponds to Clark's (2016) concept of depiction, the term we henceforth use to refer to embodied displays of music. Instead of the labels "faulted" and "prescribed" we will use "undesired" and "desired" as less normative concepts.

 Previous studies on contrast pairs in musical settings (Weeks 1996, Messner 2020) and the same phenomenon in dance classes, labelled "bodily quoting" by Keevallik (2010), suggest several key characteristics for instructional practices of this kind: the context in which they can succeed is set both by the preceding performance as well as the verbal instructions given. Some sort of verbal framing or explanation, therefore, is deemed necessary to guarantee the salience of the contrast in depictions for mutual understanding (Keevallik 2010:408). While the order of the depictions of the undesired and desired version seems to be variable, both Weeks (1996:247) and Keevallik (2010:408) identify the negative preceding the positive as the dominant pattern in their data.

 The sequential order of the components of contrast pairs marks one of the aspects under scrutiny with regard to the data set of the current contribution. Although the studies mentioned above do deal with the placement of verbal instructions in relation to depictive components of contrast pairs, big potential remains in setting up a more detailed analysis of the ways in which they are sequentially integrated. Also in need of further investigation are possible patterns which might emerge in the multimodal make-up of single components relating to their sequential placement. Furthermore, contrast is marked across depictions

through "differences in their internal performative features" (Weeks 1996:270), both in auditory and visual semiotic resources. In addition to existing analyses of vocalisations and bodily conduct during contrastive instruction, this study sheds light on both the multimodal design of instructions and on the ways in which contrast is marked in these specific instances of instruction, thus feeding into the study of multimodal stance-taking in general.

## 3.    Research questions, data and method

The aim of this case study is to disentangle the verbal and depictive components of contrast pairs, guided by the following questions: 1) How are the components of contrast pairs integrated sequentially? 2) Which combinations of semiotic resources are used to mark contrast?

To answer these, we adopt a corpus-based approach, and analyse video recordings of five conductors collected during rehearsal with their respective wind and brass orchestras in Flanders (Schrooten & Feyaerts 2020). As an exploratory selection, the first rehearsal per conductor was chosen for this study, amounting to approximately ten hours of data. Figure 1 shows the position of the orchestra and the conductor as well as the setup of camera and microphone during recording. The camera was placed behind the musicians in the last row so that the conductor is the only person fully visible in the camera frame. An additional microphone was placed on the stand in front of the conductor.
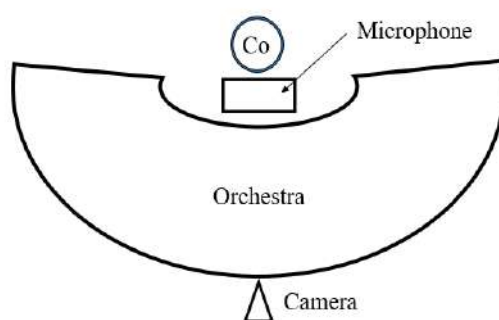


Figure 1. Recording setup during data collection

Using the annotation software ELAN (Wittenburg *et al.* 2006), 99 cases of contrast pairs were manually identified and segmented in the videos, the central

criterion for the identification of these instances being that at least one positive and one negative depiction occur in close interactional proximity to each other, adhering to the definition of contrast pairs proposed by Weeks (1996). The polarity of components was determined based on the verbal framing, context information and was discussed among authors. Single positive or negative depictions were not included, also when they are juxtaposed with verbal utterances marked with the opposite polarity. Notably, the frequency of this phenomenon differs for each of the conductors, which might be accounted for by different factors such as conducting style and expressivity, their education, their form that day, the repertoire rehearsed, and the duration of each rehearsal.

For all selected cases, we transcribed the conductors' speech at the level of intonation units to capture the focus of the participants' consciousness (Chafe 1994:63) as is common practice in Interactional Linguistics and Cognitive Linguistics (cf. Bressem 2021). Non-lexical vocalisations were annotated in line with orthographic rules representing Dutch pronunciation. To mark the order of different components of contrast pairs, we noted their polarity: positive (+), negative (-), or neutral *(N)* such as purely localising a fragment in the music, as well as their primary source of information: verbal (*v*) or depictive (*d*). Table 1 contains an overview of the different parameters:

Table 1. Annotation parameters for components in contrast pairs

| Parameter | Meaning |
|-----------|---------|
| v+ | verbal positive |
| v- | verbal negative |
| vN | verbal neutral |
| d+ | depictive positive |
| d- | depictive negative |
| dN | depictive neutral |

Thus, a depiction showing how a passage should be performed was indicated as *d+*, a verbal utterance describing how something should *not* sound, or announcing a negative depiction was marked *v-*. Localisation of fragments was annotated as *vN* or *dN*. These categories will be exemplified with data in the next section. Other aspects included in the analysis were manual gesture, facial expressions as well as body and head movement. So far, these have been annotated for the examples discussed below; a systematic annotation for all cases will follow.

## 4.    Observations

In the following close reading of two examples, we first focus on the sequential order of their components, and then on the multimodal complexity they display. The cases presented below were chosen as representative instances of more and less sequentially complex structure and different multimodal design.

### 4.1.  *Sequential complexity*

The first aspect under scrutiny is the temporal integration of components that are included in a contrast pair. Our main interest here lies in the variable order of the positive and negative polarity (in the examples discussed here, there are no purely localising, neutral components) as well as the primary source of information used, i.e. speech or depiction.

Previous studies on contrast pairs sketch certain tendencies (Weeks 1996, Keevallik 2010) suggesting a relatively straight-forward structure of two, maybe verbally framed, juxtaposed depictions, where the negative pole is more likely to precede the positive one. However, the data analysed in this study suggest a high level of variety in sequential order. Let us first look at an example that displays a relatively simple structure, which might qualify as a somewhat prototypical contrast pair. The transcriptions that follow are based on Mondada (2019) to allow for a reader-friendly presentation.[1]

```
01    nu hoor ik nu hoor ik de voorlaatste mate *&∞<<cresc>#babababababap>*&∞

      now I'm hearing now I'm hearing the second to last bars babababababap
                                           *left fist forward slowly*
                                           &head towards shoulders &
                                           ∞eyebrows raised--------∞
                                                #fig.1
```



---

[1] A turn-based annotation was avoided in first instance since the setting of orchestra rehearsals is peculiar in the way that turns by conductors would often form unfeasibly long sequences. Intonation units were the initial form of annotation in the corpus. The convention suggested by Mondada (2019) was used to present the examples here to integrate other semiotic resources and allow for a reader-friendly presentation.

```
02      ma ik moet echt hebbe *÷∞#BABABABA#BABABA*÷∞
        but I really need to have BABABABABABABA
                              *left fist beats forward seven times*
                              ÷shoulders towards head÷
                               ∞eyebrows lowered∞
                                #fig.2  #fig.3
```
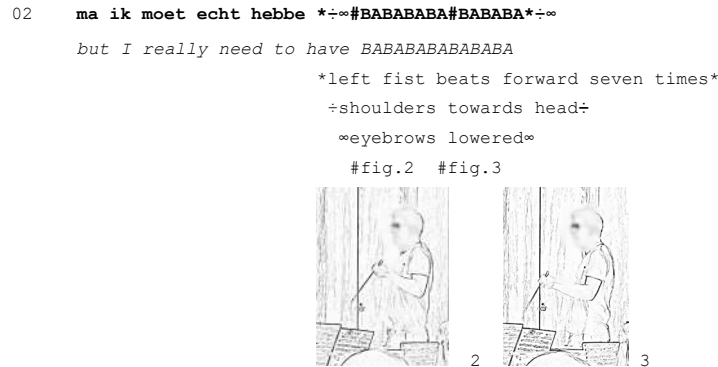


Figure 2. Example 1 – Conductor 5

Just before this first example, conductor C5 has closed a playing sequence. With the contrast pair that follows, he refers to the intensity with which a certain musical phrase just before the end of the piece should be performed. The fragment can be seen as divided into two parts: first the negative pole, second the positive one. In line 01, C5 refers to how the musicians have just played by saying "now I'm hearing the second to last bars". An embedded depiction (Hsu *et al.* 2021) follows, in which C5 re-enacts and at the same time evaluates what he heard by vocalising along with specific bodily conduct, displaying how it should not be performed. The multimodal design of this and all other depictions will be discussed in more detail in section 4.2. In line 02, a polarity switch is marked by the utterance "but I really need to have", initiating a prescription, which is again followed by a depiction of the passage in question, this time showing what the conductor in fact wants it to sound like. The fragment thus consists of two intonation units, line 01, which contains a verbal negative (*v-*) and an embedded depictive negative component (*d-*), and line 02, which qualifies as a verbal positive (*v+*) and an embedded depictive positive (*d+*).

This case thus fits what one might consider a prototypical contrast pair, containing one pair of verbally framed depictions, the negative preceding the positive part. However, the instances identified in our corpus comprise between two and nine components, displaying different orders of negative and positive parts.

Take the following example of conductor C4, who addresses the clarinet section after just having closed the play-through of an entire piece.

```
01     kunnen we *euhm* ÷als we die die prikjes÷ doen he,
       can we uh when we do these these stings eh
                 *lifts up score*
                        ÷steps towards clarinet section÷
02     #*÷<<dim>BIdeldidel#dideldi>÷*#  *÷<<dim>BIdeldideldideldi>÷*
        *right hand moves forward      *right hand moves forward
                      and backward*                   and backward*
        ÷torso sways forward           ÷torso sways forward
                      and backward÷                  and backward÷

       #fig.1              #fig.2      #fig.3
```



```
03     *diminuendo denken?*
       think diminuendo
       *right hand backwards*
04     en niet #*÷∞BIDELDIDELDIDELDIDELDI &BIDEL#DIDELDIDELDIDELDI&*÷
       and not BIDELDIDELDIDELDIDELDI BIDELDIDELDIDELDIDELDI
                  *both hands shake repeatedly---------------------*
                   ÷torso tense-------------------------------------÷
                                     &head shakes repeatedly&
                   ∞frown----------------------------------------->
                   #fig.4                        #fig.5
```



```
05     ÷(1.6)∞÷
       ----->∞
       ÷turns right and steps back to stand÷
06     geen wekker *die 's morgens afgaat ÷alstublieft.*
       no alarm clock that goes off in the morning please
                      *puts score on stand---------------*
                                      ÷torso slightly forward-->>
```

Figure 3. Example 2 – Conductor 4

Lines 01-03 represent the positive part of the contrast pair, while lines 04-06 mark the negative one. The poles each contain two verbal utterances and one depiction. In line 01, the conductor formulates a question "can we uh when we do these stings eh" with which he both localises the musical passage in question, labelled by him as "little stings", and initiates a prescriptive instruction ( *v*+). At the same time, he picks up the score from his stand and takes a step towards the musicians addressed. This is followed by a depiction (*d*+) in line 02, showing what the passage in question should sound like. As if completing the question from line 01, in line 03 he states "think diminuendo?" (*v*+), reiterating verbally what he has just depicted. Line 04 marks the switch in polarity with the utterance "and not" (*v*-) after which a second depiction (*d*-) is offered, this time displaying a previous, undesired interpretation of the "little stings" passage and an evaluation thereof. After a pause during which the conductor returns to his stand in line 05, he closes the contrast pair verbally in line 06, warning the musicians that the passage in question should not sound like an alarm clock (*v*-). This example thus consists of six components, two verbal positives (*v*+), one depictive positive (*d*+), two verbal negatives (*v*-) and one depictive negative (*d*-), spread across five intonation units.

Next to cases that more or less match the temporal order of the two examples just discussed, our data also reveal instances which contain more than one switch in polarity, similar to an example discussed in Keevallik (2010:412–413). Additionally, while the majority of cases seem to include verbal frames that index whether a depiction is negative or positive, sometimes there only is a verbal frame for one of them or none whatsoever. To gain further understanding thereof, in the course of this case study, a more detailed distinction of verbal components in contrast pairs is planned along with identifying their relation to different types of depiction (Clark 2016, Hsu *et al.* 2021).

### 4.2.  *Multimodal complexity*

We now zoom in on the multimodal complexity of the two previous examples and the way in which the contrast between undesired and desired versions is marked.

Interestingly, in example 1, while the polarity of the two components is verbally marked, the learnable, which in this case is the intensity with which a certain passage is to be played, is not made explicit through speech. Therefore, the musicians need to infer from the multimodal depictions what it is exactly they need to adjust in their play. In line 01, the utterance "now I'm hearing" indicates the negative polarity, since an elaborated reference to a previous performance with a positive evaluation is highly uncommon in this setting. When

looking at the depiction (*d*-) in line 01, we can observe the following: the conductor vocalises the syllables "babababababap" starting at a low intensity and increasing in loudness (crescendo) towards the end. He also rises in pitch, mirroring the melody in the musical phrase in question. Along with the crescendo in the vocalisation, he moves his left fist away from his body in one continuous movement on the sagittal axis. For the duration of this movement, he pulls his head towards his shoulders and raises his eyebrows, displaying a distanced and less committed stance (Streeck 2009:189–190) as if questioning or doubting. After announcing the positive part of the contrast pair in line 02 "but I really need to have", the conductor launches the positive depiction (*d*+). He vocalises the same syllable "ba" but repeats it seven times, instead of the five times before, which renders this depiction longer than the first one, thus highlighting it as the desired performance. The vocalisation starts at a higher volume than the end of the previous one and adding continued stress on each syllable. When looking at the conductor's left hand, we can again observe his fist moving along a sagittal axis, but this time in a movement repeated seven times along with the stresses in the vocalisation. These movements are more intense than the previous one in speed and tension, as well as involvement of shoulder and torso. The conductor's eyebrows are lowered, his face displaying a slight frown. The use of lowered eyebrows and pressed lips for depicting higher musical intensity (Poggi 2017:45) has previously been attributed to the conceptual metaphor of LOUDER IS LIKE BEING ANGRY (Opazo 2018:70). While in the *d*- part, his head moved down toward his shoulders, in the *d*+ it is rather the left shoulder which approaches the head during the more intense phase of the gesture. Thus, we can infer from the differences between the depictions that the learnable at stake is indeed connected to intensity and volume of the musical phrase.

In the second example, C4 also launches the contrast pair with a verbal intro (*v*+) in line 01, indexing the desired version by asking "can we", which is then depicted (*d*+) in line 02 with following qualities: the conductor vocalises the phrase labelled as "little stings" with the syllables "bideldideldideldideldi" twice in a row, with an accent on the first syllable "bi", followed by a decrease in volume (diminuendo) respectively. Along with the accent in the vocalisation, he moves his right hand sagittally away from his body and slightly downward and extends his index and middle finger at the final position. During the subsequent syllables he retrieves his hand towards his shoulder while closing his hand around the baton. In the same rhythm, his torso sways forward and backward, completing the notion of accent and retrieval across modalities. This is verbally reiterated in line 03 with "think diminuendo?" (*v*+). Thus, the learnable is made verbally explicit, accompanied by a more subtle retrieving motion of his right hand. In line 04, a depiction of the undesired version is embedded after the

verbal announcement "and not" (*v*-), which marks the contrast. This depiction (*d-*) shows qualities radically different from the first one: while the same syllables are vocalised, they are so at a significantly higher and constant volume with continued stress on each syllable. Both the conductor's hands, the left one holding the score, the right one holding the baton, as well as his head are moved in rapid repeated beats, as if shaking, while his torso is tense and his face displays a frown, indexing intensity as in example 1. After a pause in line 05, he summarises the undesired features of the depiction using verbal imagery (*v*-) in line 06 comparing the musicians' previous performance to the sound of an alarm clock. Like in the example above, the contrast pair described here concerns the intensity and volume with which a musical phrase is to be performed.

In both cases, the depictions of the less intense versions are realised with quieter vocalisations (relative to the respective contrasted ones), smooth, continuous movement and raised eyebrows. In contrast, along with the louder vocalisations, quicker and repeated movement is used to refer to more intense realisations of the passages in question. Manual gestures are combined with head movement, frowns and more tension in the body. Notably, the more intense depiction in example 2, where it marks the negative part (*d-*) shows an exaggerated, jocular mocking quality (Haugh 2014) which has also been observed by Weeks (1996:274) for negative depictions.

## 5.   Summary and outlook

With regard to the aspects of sequentiality and multimodality of the contrast pairs analysed above, we can observe that the order of undesired (-) and desired (+) components is variable. However, in line with observations from previous studies (Keevallik 2010, Messner 2020, Weeks 1996), our data reveal a tendency towards the negative version preceding the positive one. Still, we want to dig deeper into cases with a rather atypical order to identify possible patterns within them. Additionally, what appears to be more complex than expected at first is the order in which verbal and depictive parts accompany each other, since conductors switch between foreshadowing and/or summarising depictions using speech for both poles of the pair, only one or none whatsoever. This aspect is directly linked to the way in which contrastive instructions, hence stances, are constructed multimodally. Conductors tap into a wide range of semiotic resources: in the examples above, these are speech, non-lexical vocalisations, manual gesture, torso and head movement as well as facial expressions. These are used both to express layered meanings simultaneously within one depiction

but also sequentially across contrastive depictions, which can be referred to as stance-stacking (Dancygier & Sweetser 2012).

Out of this exploratory analysis, several follow-up questions have emerged. In order to gain a better understanding of action patterns that occur within contrast pairs, we need to expand the analysis along several routes. First of all, we want to further scrutinise the relation between verbal and depictive parts. To that end, we will tease apart the verbal frames within contrast pairs in greater detail to see how evaluations, instructions, identification of learnables, localisation of musical passages etc. are construed verbally. This way we can distinguish between cases like example 1, where only polarity is marked in speech but all relevant information about the qualities of undesired and desired versions lies in the depictions, and cases like example 2, where a verbal summary of the depictions is offered. Secondly, regarding the depictive components, we will identify speech-depiction relations as well as depiction-depiction relations (Hsu *et al.* 2021) to account for the role of multimodal conduct in the expression of contrastive instructions. To that end we will annotate the following semiotic resources based on the M3D annotation scheme for gesture (Rohrer *et al.* 2020) and Bressem (2021): manual gesture, head movement, torso movement, and facial expressions. Thirdly, for cases in which verbal utterances are indeed used to formulate evaluations and prescriptions, we want to indicate whether there are conceptual links across several intonation units to determine the degree of repetition or novelty of information. Additionally, we will scrutinise different instruction formats, such as questions, directives or generic statements and the choice of personal pronouns to see if any patterns emerge in relation to the depictions performed. At the same time, we hope to gain a better insight into the different viewpoints (Dancygier & Sweetser 2012) conductors assume during instruction, which is also inherently linked to the way in which their stances take form. The findings of this expanded study will flow into a typology and more in-depth study of contrast pairs in this dataset.

Concluding the close reading of two authentic corpus examples, it appears that the analysis of the ways in which conductors make use of contrast pairs offers valuable grounds for the study of instructional communication and multimodal stance-taking. Due to their evaluative and contrastive nature, the cases under scrutiny allow analysts to pinpoint exactly where and how contrast is expressed. At the same time, they let us map out the complex interplay of different markers for the expression of evaluations and prescriptions. Departing from the analysis of directly contrasted embodied instructions, which provide us with more information about learnables in question and the adjustments wished for, a wider variety of instructive actions, for instance in the form of single positive or negative depictions, can be studied.

## 6.   References

Bohmann, A., and Ahlers, W. (2021). 'Stance in narration: Finding structure in complex sociolinguistic variation'. *Journal of Sociolinguistics*, 26(1), 65–83.

Borràs-Comes, J., Kiagia, E., and Prieto, P. (2019). 'Epistemic intonation and epistemic gesture are mutually co-expressive: Empirical results from two intonation-gesture matching tasks'. *Journal of Pragmatics*, 150, 39–52.

Bressem, J. (2021). *Repetitions in Gesture: A Cognitive-Linguistic and Usage-Based Perspective*, Berlin/Boston: De Gruyter, Inc.

Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*, Chicago, IL: University of Chicago Press.

Chindamo, M., Allwood, J., and Ahlsén, E. (2012). 'Some Suggestions for the Study of Stance in Communication'. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 617–622.

Clark, H. H. (2016). 'Depicting as a method of communication'. *Psychological Review*, 123(3), 324–347.

Clift, R. (2021). 'Embodiment in Dissent: The Eye Roll as an Interactional Practice'. *Research on Language and Social Interaction*, 54(3), 261–276.

Dancygier, B., and Sweetser, E. (Eds.). (2012). *Viewpoint in Language: A Multimodal Perspective*, Cambridge University Press.

de Vries, C., Oben, B., and Brône, G. (2021). 'Exploring the role of the body in communicating ironic stance'. *Languages and Modalities*, 1(1), 65–80.

Debras, C. (2015). 'Stance-Taking Functions of Multimodal Constructed Dialogue during Spoken Interaction'. In G. Ferré and M. Tutton, eds., *Gesture and Speech in Interaction - 4th edition (GESPIN 4)*, Nantes, France: HAL, 95–100.

Debras, C. (2017). 'The shrug: Forms and meanings of a compound enactment'. *Gesture*, 16(1), 1–34.

Deppermann, A. (2018). 'Instruction practices in German driving lessons: Differential uses of declaratives and imperatives'. *International Journal of Applied Linguistics*, 28(2), 265–282.

Du Bois, J. W. (2007). 'The Stance Triangle'. In R. Englebretson, ed., *Stancetaking in discourse: subjectivity, evaluation, interaction*, Amsterdam, Philadelphia: John Benjamins Publishing Company, 139–182.

Du Bois, J. W., and Kärkkäinen, E. (2012). 'Taking a stance on emotion: affect, sequence, and intersubjectivity in dialogic interaction'. *Text & Talk*, 32(4), 433–451.

Ehmer, O. (2013). 'Veranschaulichungsverfahren im Gespräch'. In K. Birkner and O. Ehmer, eds., *Veranschaulichungsverfahren im Gespräch*, Mannheim: Verlag für Gesprächsforschung, 2–17.

Englebretson, R. (Ed.). (2007). *Stancetaking in discourse: subjectivity, evaluation, interaction*, Amsterdam. Philadelphia: John Benjamins Publishing Company.

Haugh, M. (2014). 'Jocular Mockery as Interactional Practice in Everyday Anglo-Australian Conversation'. *Australian Journal of Linguistics*, 34(1), 76–99.

Heritage, J. (2012). 'The Epistemic Engine: Sequence Organization and Territories of Knowledge'. *Research on Language and Social Interaction*, 45(1), 30–52.

Hsu, H.-C., Brône, G., and Feyaerts, K. (2021). 'When Gesture "Takes Over": Speech-Embedded Nonverbal Depictions in Multimodal Interaction'. *Frontiers in Psychology*, 11, 1–23.

Ivaldi, A., Sanderson, A., Hall, G., and Forrester, M. (2021). 'Learning to perform: A conversation analytic systematic review of learning and teaching practices in performing arts lesson interactions'. *Learning, Culture and Social Interaction*, 28, 1–17.

Jehoul, A., Brône, G., and Feyaerts, K. (2017). 'The shrug as marker of obviousness'. *Linguistics Vanguard*, 3(1), 1–9.

Kärkkäinen, E. (2007). 'The role of I guess in conversational stancetaking'. In R. Englebretson, ed., *Stancetaking in discourse: subjectivity, evaluation, interaction*, Amsterdam, Philadelphia: John Benjamins Publishing Company, 183–219.

Keevallik, L. (2010). 'Bodily Quoting in Dance Correction'. *Research on Language and Social Interaction*, 43(4), 401–426.

Meissl, K., Feyaerts, K., and Sambre, P. (2022). 'Mapping musical dynamics in space. A qualitative analysis of conductors' movements in orchestra rehearsals'. *Frontiers in Communication*. doi:10.3389/fcomm.2022.986733

Messner, M. (2020). 'Gesangliche Demonstrationen als instruktive Praktik in der Orchesterprobe'. *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion*, 21, 309–345.

Mondada, L. (2019). 'Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction'. *Journal of Pragmatics*, 145, 47–62.

Opazo, P. (2018). *The moving and expressive body: A study on the semiotic resources used by Classical Orchestra Conductors and their metaphorical associations in the presentation of musical dynamics.*, Vrije Universiteit Amsterdam, Amsterdam.

Poggi, I. (2017). 'Signals of intensification and attenuation in orchestra and choir conduction'. *Normas*, 7(1), 33.

Reed, D., and Szczepek Reed, B. (2014). 'The emergence of learnables in music masterclasses'. *Social Semiotics*, 24(4), 446–467.

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., … Prieto, P. (2020). 'The MultiModal MultiDimensional (M3D) labelling scheme for the annotation of audiovisual corpora', Presented at the Gesture and Speech in Interaction (GESPIN), Stockholm, Sweden.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). 'The preference for self-correction in the organization of repair in conversation'. *Language*, 53(2), 361–382.

Schrooten, S., and Feyaerts, K. (2020). 'Conducting Fanfare Orchestras. A multimodal corpus.', KU Leuven MIDI.

Sinclair, J. M., and Brazil, D. (1982). *Teacher talk*, Oxford: Oxford University Press.

Stivers, T. (2008). 'Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation'. *Research on Language and Social Interaction*, 41(1), 31–57.

Stoeckl, H., and Messner, M. (2021). 'Tam pam pam pam and mi – fa – sol: constituting musical instructions through multimodal interaction in orchestra rehearsals'. *Multimodal Communication*, 10(3), 193–209.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*, Vol. 2, Amsterdam, Philadelphia: John Benjamins Publishing Company. doi:10.1075/gs.2

Weeks, P. (1996). 'A Rehearsal of a Beethoven Passage: An Analysis of Correction Talk'. *Research on Language and Social Interaction*, 29(3), 247–290.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). 'ELAN: a professional framework for multimodality research'. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, 1556–1559.

Zemel, A., and Koschmann, T. (2014). '"Put your fingers right in here": Learnability and instructed experience'. *Discourse Studies*, 16(2), 163–183.