Tim Ihde

Die da!?! Potentials of Pointing in Multimodal Contexts

Abstract

The following article focuses on complex pointing structures in multimodal contexts in the German music video *Die da!?!* by Die Fantastischen Vier. Departing from the two-field-theory of Karl Bühler and the concept of multimodality by Ellen Fricke, these structures are described. ›Pointing words‹ (deictics) and nonverbal pointing structures in combination change the qualities of their references. The potentials and problems of multimodal pointing complexes, which emerge through the interaction of the visual level, sounds, and editing, are different from these of their elements. This article describes how these complexes work and that they are more than just the sum of their parts.

1. Introduction

To refer to a point is a natural form of communication of human culture. Pointing is quite easily realizable by using the hand or the gaze and even words are able to refer to a point, albeit, some prerequisites need to be fulfilled. Verbal pointing in combination with other pointing elements, such as hands or gaze, can lead to complex structures that have the possibility to affect each other mutually. So far, the additional value of these structures was seldom in the focus of linguistic research. But with the growing awareness of the multimodality of language, complex pointing structures were given more and more prominence (see FRICKE 2012). The discoveries out of the research of these multimodal complexes are an argument for the research on multi-

modality in general. With a perspective on multimodality, there are several original aspects focused that are not included in the research on monomodal phenomenons (see chapter 4.2).

The following article discusses complex multimodal pointing structures occurring in the German music video *Die da!?!* by the hip hop formation Die Fantastischen Vier from the early nineteen-nineties.[1] In this music video, deictic forms are highly frequent and, moreover, assume specific qualities that result from the simultaneous combination of the auditory and verbal sphere of the music video.

In a first step, the theoretical fundamentals of linguistic pointing theory are explained departing from Karl Bühler's two-field-theory. In a second step, the concepts of media and multimodality are discussed, for the purpose of introducing necessary terminology. In this connection, it is important to distinguish between multimediality and multimodality as done in section 3. Afterwards, the potentials of pointing structures within the music video are discussed by showing special cases of multimodal pointing complexes.

2. Deixis Theory

The psychologist Karl Bühler published his *Theory of Language* (*Sprachtheorie*) in the mid nineteen thirties, which, since then, had been reprinted several times without major changes. Apart from many others, one basic innovative thought of Bühler's theory is the differentiation between two word categories: naming words and pointing words or so-called deictic words (cf. BÜHLER 1990: 121). After Bühlers concept, naming words are located in a symbolic field, while deictic words are said to appear in a deictic field. Accordingly, Bühler calls this the »*two-field theory*« (BÜHLER 1990: 95; original emphasis).

The meaning of naming words is mostly conventionalized through social semantic negotiations and therefore understandable without further specification through the established situation. Words like ›heart‹, ›sing‹ and ›green‹ belong to the symbolic field caused by their kind of referencing. Naming words are symbols in the sense of Charles Sanders Peirce (cf. PEIRCE 1998: 65).

However, deictic words like ›here‹, ›then‹ or ›you‹ change their meaning depending on the situation in which they are used. For example, the reference of ›I‹ varies every time the speaking person changes. Accordingly, a deictic word (and every other kind of deictic expression) is not fully interpretable without further information. For example, if the title screen of a movie only shows the adverb ›later‹ as a temporal deixis, it will inevitably evoke some semantic disturbance leading to the question ›later from what?‹. Therefore, every deictic form requires at least a related starting point in order to be interpreted. Bühler calls this »coordinate source« (BÜHLER 1990: 117) of the

---

[1] Die Fantastischen Vier: *Die da!?!* (1992). Hamburg: EMI Music Publishing Germany GmbH.

deictic field the »origo« (BÜHLER 1999: 102). Out of the origo, every other point in time and space can be referred (cf. BÜHLER 1999: 107) and even the origo itself can be the center of a deictic reference by the usage of the verbal forms ›here‹, ›I‹, and ›now‹:

> Let two perpendiculary intersecting lines on the paper suggest a coordinate system to us, 0 for the origin, the coordinate source[.] [...] My claim is that if this arrangement is to represent the deictic field of human language, three deictic words must be placed where the 0 is, namely the deictic words *here*, *now* and *I*. (BÜHLER 1990: 117; original emphases)

Without the relation to an origo, no deicticity can be established. Thus, the relation to an origo, as a specific case of context relation, is the conditio sine qua non for deicticity (cf. BÜHLER 1990: 122; BÜHLER 1999: 107; FRICKE 2007: 53). Furthermore, the origo-relation is differentiated by the question: is the origo part of the reference or not? The reference of deictics like ›we‹, ›here‹ and ›now‹ include the origo and are therefore origo-inclusive. In the case of exclusion, they are called origo-exclusive (cf. FRICKE 2007: 92–93). There are even deictics that are not marked as to whether they include the origo or not. The German ›da‹, for instance, is unmarked pertained to the inclusion of the origo.

In addition to origo in- and exclusion and the three dimensions, deictics can be classified according to the type of the reference object. The referred object of verbal deictics like ›there‹, ›then‹ and ›we‹ have no clear shape and are therefore called areal deixis (›Bereichsdeixis‹). However, words like ›that‹, ›tomorrow‹ or ›I‹ are more precise by pointing to an entity. These are classified as entity deictics (›Entitätendeixis‹) (cf. FRICKE 2007: 99). Besides, pointing gestures can also be differentiated in relation to their type of reference. For example, by pointing with the index finger, a special point is being addressed, why this kind is defined as spatial point deixis. Pointing with the palm, otherwise, is declared to be a directional deixis (›Richtungsdeixis‹) (cf. FRICKE 2007: 99). The enlisted typification of deictics is quite more profound, but the given outline is adequate for the subsequent thoughts on the potentials of multimodal deictic complexes in the music video *Die da!?!*. However, the additional definition of multimodality is indispensable.

3. Multimediality and Multimodality

The awareness of multimodal aspects was a gradual development in linguistic research. At first, isolated examples of (written) verbal utterances stood in the focus of linguistics. Thus, the influence of a medium on oral or written expressions had been unnoticed for a long time span. The inclusion of medial aspects in linguistic studies is rather new. It is today widespread beyond disciplinary borders that mediality is a non-negligible foundation of communication. Complex communicative structures often involve multiple media which sets the term ›multimediality‹ relevant. To outline the differences between

multimediality and multimodality, it is necessary to explain the basical concepts behind the complex word formations. First, ›medium‹ is a pervasive term with countless different definitions. In the eighties, Posner differentiated between six types of media-terms: biological, physical, technological, sociological, cultural related and code or sign system related (cf. POSNER 1986: 293–303). All media are systems of the production, distribution and reception of signs, which influence the generated sign processes in a known and constant way (cf. POSNER 1986: 293).

Biological media are distinguished based on the perception organs they address. In this sense, there are visual, auditory, tactile, olfactory and gustatory media (cf. POSNER 1986: 293–294). Hence, code or sign system related media are categorized based on how sender and receiver link sign carriers and messages (cf. POSNER 1986: 296). Thereby, code is understood as a semiotic system for the usage of signs (cf. NÖTH 2000: 219).

Since Posner's account, several other distinctions of media have been proposed. The term ›multimedia‹ is based on the assumption that two or more media are involved in one specific situation addressing at least two different perception channels (cf. POSNER 1986: 300; FRICKE 2012: 46). We do not go into further detail here because for the present paper only the biological and the code related terms of media by Posner are relevant and build the theoretical base for multimodality.[2]

Like for the term ›media‹, there are also several definitions of ›multimodality‹. They open up a continuum between the understanding of multimodality as an omnipresent phenomenon and as a special case of semiotic relation within media. The second notion is preferred in this article. Following Fricke, we understand ›multimodality‹ in the narrower sense:

> If two linguistic media are structurally and/or functionally integrated into the same code at the same time, or if, conversely, one code manifests itself simultaneously in two different media, then we can speak of multimodality. (FRICKE 2013: 739)

Different from multimodality, ›multimediality‹ does not necessarily mean that there is a mutual relation between the involved elements. The codes in a multimodal relation are more than just present at the same time. They are semantically interwoven with each other, resulting in a special form of meaning negotiating process, subsumed in the notion of ›multimodality‹.

---

[2] For a more detailed discussion of multimodality see SCHNEIDER/STÖCKL 2011.

4. »Die da«—a Deictic Complex

4.1 Preliminary Considerations

Die da!?! was the first commercial successful German hip hop track published in the early nineteen-nineties. It is mainly a fictitious dialogue between two persons (Smudo and Thomas D.) that both met a woman at the weekend and tell the other what happened since then. The two women, that they are talking about, are absent from the talk what causes Smudo as well as Thomas D. to try to locate them in their perceivable space by pointing on different women with the verbal deictic complex »die da«. After repeated attempts without success, one of the described women joins the room. Her occurrence causes them to realize that they are both talking about the same person that was cheating on both of them yet with another man.

›Die da‹ is by far the most frequently used verbal deictic combination occurring in the chorus of the rap song, which is even named after it. Because there is no equivalent structure in English Language, its functionality has to be described first by focusing on its elements. Although, many different references can be established with ›die‹, in this specific case, it points to a female person. Therefore, it is situated in the personal deictic dimension. By following the subclassification of Fricke (see chapter 2), ›die‹ is an entity deixis. ›Da‹, however, might be translated with ›there‹ and points to an unspecified local area. Therefore, it is an areal deixis, operating in the local deictic dimension. Different from ›die‹, ›da‹ is not specific in marking whether it includes or excludes the origo. Both options are possible and imaginable.

The combination of ›die‹ and ›da‹ leads to a verbal deictic complex that references a concrete point in an area, where the person is located. Through the addition of another element, the created deictic complex is specified by involving two deictic dimensions. Nevertheless, even the specification of the complex is not sufficient to locate the referred point, without the addition of another deictic quality, neither in the room nor in the imagination. The addition through a nonverbal deictic clue (›Zeighilfe‹) (cf. BÜHLER 1990: 141), which gives a directional information, seems necessary. To indicate this direction, numerous types of nonverbal pointing are used, e.g., pointing with gaze, lips, nose, the whole hand or the fingers (see FRICKE 2014: 1804).

The additional information offered by the visual sphere of the music video provides several different deictic clues. For the appearance of ›die da‹, it has to be differentiated between a deictic complex without any nonverbal pointing, which is a spatial point deixis pointing at a female person within an area and a deictic complex with the simultaneous use of nonverbal pointing, which also refers to a female person within an area, but is even more specific, by giving additional information on the direction which leads to the creation of a multimodal deictic complex.

4.2 Multimodal Complexity

As mentioned in chapter 3, multimodality is a special case of a mutual media relation. One aspect of setting music videos relevant for multimodal research is the additionality of their visual sphere. The codes of a music video do not necessarily stand in a multimodal relation. The auditory sphere is often produced before any idea of an additional video track is created. Therefore, intended mutuality cannot be assumed. For this reason, multimodality in a music video is a temporal phenomenon and not ubiquitous. For example, the visual and the verbal spheres shown in figure 1 do not correlate nor do they create a multimodal medial complex, albeit they appear at the same time. The depicted person is not even the referred object of the verbal deictic complex ›die da‹. The simultaneous presence of the visual and auditory information seems to be arbitrary, because the involved codes are not interacting with each other.

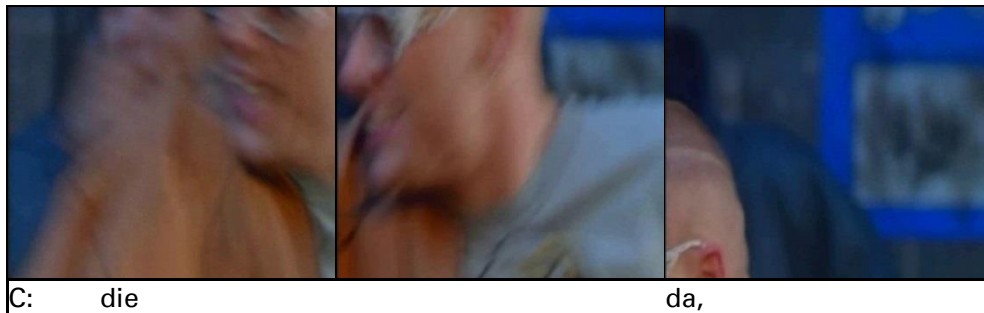C: die   da,

Fig. 1

However, the deictic complex of figure 2 shows a multimodal deictic complex by the simultaneous occurrence of the visual perceivable pointing gesture and the verbal entity deixis ›die‹. The first part of the directional deixis relates to the personal dimension by using »die« as an entity deixis and the local dimension by showing the arm corresponding with the index finger and the gaze direction. The person indicating with arm, hand and gaze is functioning as the origo of the deictic complex, yet the articulation movement of ›die‹ in the picture sequence seems to be bound to the audible word, albeit the depicted distance of the sound source and the actual sound presence do not fit.

Fig. 2

The image area that is involved in the visual deixis is incremented by the slight rotation of the head in the direction of the arm. Thus, according to Bühler, the appealing function of the image sequence is amplified. The presented deictic complex is pointing to a room point bound to a personal entity, but due to the mediality of the image frame, there is no information about the concrete reference object at this point of time.

In the following frame of the next shot, however, a referenced object is offered by presenting a female person.

Fig. 3

The functionality of ›da‹ is to mark an undefined part of an area that leads to an actualization of it. With the change of the frame to figure 3 the whole pictorial area is actualized simultaneously with ›da‹. In conclusion, the marked area is the complete picture actualized by the editing of the video sequence. Therefore, the new picture is the area for the referenced object of the multimodal deictic complex. At this point, the reference includes a component of the personal as well as the local deictic dimension.

The cut within the fabrication of the deictic complex in combination with the knowledge of the complete deictic complex out of former repetitions creates an indexical relation. Hence, the woman of the second shot becomes the referenced object. In a succession of two shots, at first showing a pointing person and second a referable object, the first shot shows the source of the deixis (origo) and the second its object (cf. FRICKE 2015: 160). Such a sequence

tries to compensate restrictions of the medial format in comparison to a phy-logeneticly sooner established communication situation which knows language as an audio-visual medium (cf. FRICKE 2015: 148). This kind of indexicality in combination with the linearity of the video provides an additional temporal component of the deictic complex, considering film as three-dimensional, whereby the third dimension is time, not spatial depth (cf. SACHS-HOMBACH 2006: 228). In the theoretical construction of the present sequence shown in figure 2 and 3, the multimodal deictic complex creates a new quality of deicticity which is an omnidimensional deixis.

S: die   da

Fig. 4

Looking at figure 4, another phenomenon becomes evident. Taking the direction of the gaze as an additional vector of the deictic clue, the point of intersection is set relevant. Obviously, the position of the woman's head correlates with the region of the raised index finger. Thus, the spanned two-dimensional surface quality of both pictures is also used as reference region.

Yet looking at this structure, there is another interesting aspect. The German noun ›Auto‹ (car) has two synonyms that both share the feminine genus (›Karre‹ or ›Kutsche‹). Accordingly, both, the car or the woman could be the reference object of the deictic structure indicated by ›die da‹. However, with the visual information, this vague reference is disambiguated.[3]

The picture sequence of the video illustrates a movement mimetically resulting in an easily perceptible impression (cf. SACHS-HOMBACH 2006: 228). The sequenced pictures are contrasted with the following one that is used in the editing. Narrative strategies had been developed, realized by a continuous image stream. Jumps of the temporal and the local perspectives of a film advanced to regular forms of movie narration. This type of narration occurs in relation to a perspective created with the camera and leads to the identification of the visual perspective of the recipient and the camera (cf. SACHS-HOMBACH 2006: 230). From a deixis-theoretical point of view, the camera is offering another origo. The filmed events stand in relation to the camera per-

---

[3] I would like to thank Ellen Fricke for pointing this out during the preparation of my talk »Die da!?! – Potentials of Pointing in Multimodal Contexts« for the Winter School »Mediality and Multimodality across Media« in Tübingen 2015, on which this article is based.

spective and influence the attention of the viewer. Therefore, the camera has the potential to work as a deixis.

In another image sequence (fig. 5), there is a special case illustrating the additional value of the camera perspective. The deictic clues (gaze and gesture) of the person in the right part of the picture are nearly referencing the center of the image.

C:genau   die   da

Fig. 5[4]

Through the pointing into the center of the screen, the supposed viewer respectively the location of the viewer is addressed as well as the female person wearing the green safety helmet. The complex deixis is therefore bidirectional and ambiguous. In combination with the German word ›genau‹, which carries the propositional function of illustrating clarity and can be translated with ›exactly‹, this multiplex reference remains doubtful. Thus, the double entendre of this deictic complex becomes pithy.

Involving the following shot (shown in fig. 6), the familiar scheme of origo plus deixis in the first image sequence and the referenced object in the next shot is repeated.

C: es   is'   die   da   die

Fig. 6

Therefore, the deictic structure of figure 5 has three different referenced objects: (1) the center of the picture (fig. 5) which is bound to an entity, (2) the center of the sequence in figure 6 which is also bound to an entity and (3) the

---

[4] The pictures of this sequence were brightened up to increase their visual quality.

room point in front of the television reproducing the music video which is potentially bound to an entity. The multimodal deictic complex is ambiguous with regard to his reference object (fig. 7).

Fig. 7

The direct look into the camera is unusual for our viewing habits and is thus a clearly marked stylistic device. Kress and van Leeuwen draw a strict distinction between pictures that address the recipient by eye contact or gestural reference and those that do not dispose of such a kind of reference. They attribute the former type two functions:

> In the first place it creates a visual form of direct address. It acknowledges the viewers explicitly, addressing them with a visual ›you‹. In the second place it constitutes an ›image act‹. The producer uses the image to *do* something to the viewer. It is for this reason that we have called this kind of image a ›demand‹. (KRESS/VAN LEEUWEN 2010: 117–118; original emphasis)

The pointing into the camera marks the perspective as a special form of a subjective. But, this camera perspective does not represent the look of a person, it rather imposes it on to the viewer. It is an appeal to the normally passive recipient. The sole addressed person hereby stands vicariously for a disperse audience which has no homogeneous position in the reception situation. Thus, the visual offered origo does not fit to the perception of the addressees, instead, it fits to the picture reproducing medium, the screen. A kind of interface between the technical reproduced and the real word is established.

5. Conclusion

By using the music video *Die da!?!* by Die Fantastischen Vier, the article presented an analysis of complex multimodal deictic structures. The combination of pictures, verbal language, gestures, and editing, as present in the video example, creates deictic potentials that have not been described nor classified so far. This article, thus, offers a first description of this multimodal phenomenon from a linguistic point of view. By bringing together deixis, gesture, sound and moving image, the article shows the complex nature of deictic structure and highlights the importance of the particular modalities in establishing reference. Further studies investigating this phenomenon would contribute to a better understanding of the nature of multimodal deictic structures. In particular, an interdisciplinary perspective can offer new insights and address, for instance, the following questions: Which contribution can be given by other science and technologies like eye-tracking? How can discoveries be reintegrated in new instruments like virtual space or augmented reality? All things considered, these are thoughts for further research.

References

BÜHLER, KARL: *Theory of Language. The Representational Function of Language*. Translated by Donald Fraser Goodwin. Amsterdam [Benjamins] 1990

BÜHLER, KARL: *Sprachtheorie. Die Darstellungsfunktion der Sprache* (1934). Stuttgart [Lucius & Lucius] 1999

FRICKE, ELLEN: *Origo, Geste und Raum. Lokaldeixis im Deutschen*. Berlin [De Gruyter] 2007

FRICKE, ELLEN: *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken*. Berlin [De Gruyter] 2012

FRICKE, ELLEN: Towards a Unified Grammar of Gesture and Speech. A Multimodal Approach. In: MÜLLER, CORNELIA; ALAN CIENKI; ELLEN FRICKE; SILVA H. LADEWIG; DAVID MCNEILL; SEDINHA TESSENDORF (eds.): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Vol. 1. Berlin [De Gruyter Mouton] 2013, pp. 733–754

FRICKE, ELLEN: Deixis, Gesture, and Embodiment from a Linguistic Point of View. In: MÜLLER, CORNELIA; ALAN CIENKI; ELLEN FRICKE; SILVA H. LADEWIG; DAVID MCNEILL; JANA BRESSEM (eds.): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Vol. 2. Berlin [De Gruyter Mouton] 2014, pp. 1803–1823

FRICKE, ELLEN: Die (ab)geschnittene Hand in der Talkshow. Zur Fortschreibung antiker rhetorischer Traditionen in Bildwahl und Schnitt. In: GIRNTH, HEIKO; SASCHA MICHEL (eds.): *Polit-Talkshow. Interdisziplinäre Perspektiven auf ein multimodales Format*. Stuttgart [Ibidem] 2015, pp. 143–166

KRESS, GUNTHER; THEO VAN LEEUWEN: *Reading Images. The Grammar of Visual Design*. London [Routledge] 2010

NÖTH, WINFRIED: *Handbuch der Semiotik*. 2nd edition. Stuttgart [J.B. Metzler] 2000

PEIRCE, CHARLES SANDERS: *Phänomen und Logik der Zeichen*. Translated by Helmut Pape. Frankfurt/M. [Suhrkamp] 1998

POSNER, ROLAND: Zur Systematik der Beschreibung verbaler und nonverbaler Kommunikation. Semiotik als Proprädeutik der Medienanalyse. In: BOSSHARDT, HANS-GEORG (ed.): *Perspektiven auf Sprache. Interdisziplinäre Beiträge zum Gedenken an Hans Hörmann*. Berlin [De Gruyter] 1986, pp. 267–313

SACHS-HOMBACH, KLAUS: *Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft*. Köln [Halem] 2006

SCHNEIDER, JAN GEORG; HARTMUT STÖCKL (eds.): *Medientheorien und Multimodalität. Ein TV-Werbespot – Sieben methodische Beschreibungsansätze*. Köln [Halem] 2011