Social Interaction

Video-Based Studies of Human Sociality

_____

Transcribing Facial Gestures:

Combining Jefferson with the International SignWriting Alphabet (ISWA)

_____

Carolin Dix

University of Innsbruck

Abstract

Within multimodal interaction analysis, transcripts serve not only as a tool for managing the volatility of interaction and catalyzing analytic procedures, but also as an essential medium for making analytic results intersubjectively available. While there are already well-established conventions for transcribing verbal and vocal interactional resources, researchers still struggle with adequately aligning and recognizably representing visual-bodily resources. This contribution provides a practical solution for multimodal transcription, combining conventions of the Jeffersonian system with the sign inventory of the International SignWriting Alphabet (ISWA). The result is a standardized, data oriented, expandable system that relies on iconic depictions rather than on verbal descriptions of visual conduct.

Keywords: multimodal transcription, facial gestures, ISWA, Sutton SignWriting

1. Introduction

Transcripts – the written representations of verbal, vocal and visual elements of social interaction – are an essential and even mandatory tool when analyzing naturally occurring human social interaction (Wagner, 2020: 296). Within the framework of (multimodal) Conversation Analysis (CA) and Interactional Linguistics (IL), transcripts provide the means to 'describe and analyze the practices used by the participants to construct the actions and events that make up their lifeworld' (Goodwin, 2000: 166). In this respect, they are a 'fundamental empirical tool with which to document and conceptualize the temporality, projectability, sequentiality and accountability of action in social interaction' (Mondada, 2019b). Transcribing social interaction is, furthermore, considered to be a practical researcher's skill (Bezemer & Mavers, 2011: 191, Jenks, 2018: 118). As an instrument and important step in processing data of authentic interaction within Multimodal Interaction Analysis (MIA), transcripts support the whole process of analysis. They play an important role in:

- identifying relevant details while preparing video data for analysis (Luff & Heath, 2015: 373; Mondada, 2019b)

- documenting and sharing data with other researchers during data sessions (Bezemer, 2014: 155; Albert et al., 2019)

- making research results transparent and accessible to other researchers in presentations and articles (Luff & Heath, 2015: 369; Mondada, 2018: 87; Wagner, 2020: 296; Aarsand & Sparrman, 2021: 293)

- giving evidence to the analytic arguments and highlighting specific features of the analysis – thus leading the reader towards the selected phenomena (Goodwin, 2000: 163; Mondada, 2007: 811; Laurier, 2014: 244; Jenkins, 2018: 119).

Even though transcripts represent certain choices and decisions made by the researcher, such as about the data extract to be transcribed and the interactive resources captured and represented in the transcript (Bezemer, 2014: 155; Wagner, 2020: 297; Aarsand & Sparrman, 2021: 290) and inevitably reduce the complexity of an interaction, they should allow for a detailed and microanalytic investigation and understanding of the moment-by-moment construction of the phenomenon of interest (Stukenbrock, 2009: 147; Ayaß, 2015: 510; Hepburn & Bolden, 2017: 101). In this respect, the process of transcribing is not only a method for preparing the data for analysis but is already a part of the analysis.

CA, IL and MIA research has further highlighted the fact that face-to-face interaction is a multimodal event and should be analyzed as such. Consequently, verbal, vocal as well as visual resources have to be taken into account while analyzing interactants' action formation and organization of talk. The video turn

and the related multimodal turn of the past years (Knoblauch & Heath, 2006; Heath et al., 2010: 5; Mondada, 2019b) have led to a growing interest in the multimodality of interaction and to fine-grained microanalysis. However, new methodological challenges have arisen:

> The analysis of video data has enabled (or rather, obliged) researchers to address the complexity of social actions from a holistic perspective. These developments go hand in hand with the problem of representing, for readers and viewers, the phenomena that are at the core of the analysis. (De Stefanie, 2022: 8)

For the verbal part of interactions, sophisticated conventions for representing features such as the spoken word, prosody (e.g. pitch, accents, loudness, speed, etc.) and sequential aspects (e.g. overlaps) have already been developed (e.g. Jefferson, 2004). However, they do not meet the requirements for representing visual-bodily behaviour (Bezemer & Mavers, 2011: 192). Although there are several suggestions on how to include representations of visual actions into multimodal transcripts (e.g. Mondada, 2019a; Luff & Heath, 2015), numerous researchers have pointed out that there is still no uniform and standardized convention that can be adapted properly to different research questions as well as different data sets (Mittelberg, 2007: 227; Stukenbrock, 2009: 146; Bohle, 2013: 993; Luff & Heath, 2015: 367; Hepburn & Bolden, 2017; Mondada, 2018: 88; Wagner, 2020: 297). However, a standardized system would help to compare and interpret research results. The issue of how to transcribe facial configurations and movements of the mouth, the eyes, eyebrows and so on, seems likely to expand this discussion even further.

This contribution takes up the desideratum of finding a standardized and expandable system that provides solutions for the adequate representation of visual-bodily resources. To this end, the paper proposes a transcription system which combines Jefferson's (2004) well-known convention for verbal and vocal transcription with the inventory of the International SignWiting Alphabet (ISWA; Sutton, 2010).[1] Besides giving practical insights into how to use the ISWA within transcripts for MIA purposes and discussing the benefits and limitations of the introduced system in general, this paper specifically focuses on ways to transcribe the positions and movements of facial resources like the mouth or the eyebrows.

The following section outlines the basic requirements and quality features of multimodal transcripts (Section 2). Then, the system of Sutton SignWriting and the ISWA will be explained (Section 3) before introducing the new system

---

[1] The system presented in this article is not limited to the combination of Jefferson and the ISWA but is also appropriate for the combination of other systems of verbal and vocal transcription with the ISWA such as GAT2 (Selting et al., 2009; see Dix, 2021 and 2022, and Dix & Gross, 2023/this issue).

(Section 4) by means of several examples of different data sets with regard to the representation of facial movements (Section 5). Finally, this paper discusses the benefits as well as the challenges of the system introduced herein (Section 6).

2. Transcribing Visual Conduct

Ever since Ethnomethodology and Conversation Analysis researchers began to analyze video recordings of naturally occurring social interaction, they have discussed ways of transcribing verbal, vocal and visual resources (see the overview in Hepburn & Bolden, 2017: 101–103). Most of the established conventions focus primarily on one specific resource: for gaze, see Goodwin (1981) and Rossano (2013); for manual gestures, see Kendon (2004) and Streeck (2009). Other researchers have developed systems that represent different interactional resources – including facial movements (Heath et al., 2010; Luff & Heath, 2015; Mondada, 2019a).

Although the transcripts differ in layout (e.g. turn-by-turn transcript vs. score system), all authors highlight that multimodal CA transcripts should 'explore the occasioned, emergent, and sequential character of practical action and interaction' (Luff & Heath, 2015: 367) and therefore cover the 'various resources that participants draw upon in the practical and concerted accomplishment of particular activities' (Luff & Heath, 2015: 368). In this respect, transcripts representing the multimodality of interactions should fulfil essential requirements and quality features:

- **Adaptability** to different data sets, the specific research question and the purpose of the transcript (working transcript vs. presentation transcript; Goodwin, 2000: 163; Bezemer, 2014: 158; Laurier, 2014: 245; Luff & Heath, 2015: 382; Hepburn & Bolden, 2017: 101; Mondada, 2018: 88; Jenks, 2018: 125).

- **Adequacy and accuracy**: precise and detailed representation of all relevant resources and phenomena (Sager, 2001: 1069; Hepburn & Bolden, 2017: 103; Mondada, 2018: 88; Jenks, 2018: 125; Wagner, 2020: 298). In this respect, such diverse and heterogeneous visual aspects as manual gestures (movements and positions of arms and hands), facial gestures (movements and positions of lips, eyebrows, cheeks, forehead, etc.), movements and positions of the head, body orientation as well as handling objects should be included (Goodwin, 2000: 163; Luff & Heath, 2015: 369).

- **Clear alignment** of verbal and visual-bodily conduct and representation of the sequential, spatial and temporal unfolding of visual behaviour; representation of the coordination and orchestration of multiple verbal and visual actions (Goodwin, 2000: 162; Bezemer, 2014: 161; Laurier,

2014: 237; Luff & Heath, 2015: 373; Mondada, 2018: 88; Mondada, 2019a).

- **Reduction of metacommunicative statements** and verbal descriptions of visual behaviour (transcriber comments), neutral representation of relevant phenomena (Sager, 2001: 1070; Hepburn & Bolden, 2017: 108).

- **Readability and clarity** despite the growing complexity of the multimodal transcript (Sager, 2001: 1069; Hepburn & Bolden, 2017: 106; Jenks, 2018: 124; Wagner, 2020: 299).

- **Writability** of signs with standard technical tools and software (Sager, 2001: 1069; Wagner, 2020: 299).

- **Flexibility and variability** to supplement, expand or shorten already existing transcripts as well as the granularity and therefore the level of detail within the transcripts according to the research question and the purpose of the transcript (Sager, 2001: 1070; Laurier, 2014: 236; Hepburn & Bolden, 2017: 113; Jenks, 2018: 124; Mondada, 2019b).

Due to the plurality of these requirements, it is obvious that a single transcription system can hardly satisfy all of them. Some criteria (like readability and detailedness) even seem to be contradictory. Besides that, the call for avoiding or reducing meta-communicative descriptions of visual-bodily behaviour seems to be a particular challenge for designing multimodal transcripts. Even the frequently used conventions proposed by Heath (Luff & Heath, 2015) and Mondada (2019a) still use transcriber comments, even if they are kept to a minimum. Transcriber comments are not only cumbersome and impractical in terms of aligning synchronous multimodal behaviour, but they also contain the risk of premature interpretations that overlay the analytic process (Sager, 2001: 1072). This becomes even more relevant when it comes to transcribing fleeting and small facial movements.[2]

For representing the unfolding of movements as well as the multimodal gestalt of actions, most researchers add video stills to their transcripts. These 'visual representations have the advantage of being easily interpretable and more holistic in representation, giving readers an easy access to multiple modalities

---

[2] Within (emotion) psychological research, the most commonly used system to represent movements and positions of facial resources is the Facial Action Coding System (FACS) developed by Ekman et al. (2002). It is a standardized 'anatomically-based descriptive research tool' (Waller & Smith Pasqualini, 2013: 921) for a 'systematic description of the component muscle movements of facial expressions' (Waller & Smith Pasqualini, 2013: 919). Each muscle movement is coded as an Action Unit (AU). However, because FACS was developed with a very detailed physiological view on facial movements for categorizing and clustering muscle activities within the face and linking them to particular emotions, it appears to be unsuitable for MIA research interests.

involved in carrying out embodied actions' (Hepburn & Bolden, 2017: 122). Other authors include line drawings, stick figures and comic-like photo arrangements in transcripts to depict visual behaviour (Goodwin, 2000; Bezemer & Mavers, 2011: 199–202; Laurier, 2014; Hepburn & Bolden, 2017: 102; Albert et al., 2019). However, many researchers are critical when it comes to using specifically designed iconic signs (see Section 3). The aim of this contribution is to show how integrating iconic, deictic and symbolic signs (such as the International SignWriting Alphabet) into multimodal transcripts can improve transcripts – especially with regard to transcribing facial movements – while meeting the issues of readability, flexibility and adaptability.

3. Sutton SignWriting and the ISWA

Some researchers (e.g. Birdwhistell, 1973) have considered alternative ways of integrating representations of visual-bodily behaviour into transcripts – for example, using the iconic Laban system (Drew, 2017: 5f; Kennedy, 2013; Sutton, 1981/1982). Even though systems like Labanotation and Sutton SignWriting (SSW) are both particularly designed for capturing visual conduct, they have not prevailed within the transcription tradition of interaction analytic research. This is due to recurring critique, which argues that iconic signs are too complex, not flexible enough to represent conventionalized forms of dance and sign languages as well as ad hoc forms of naturally occurring social action, difficult to read, not easy to learn and difficult to produce with standard software (Sager, 2001: 1072; Luff & Heath, 2015: 368).

In the 1970s and 1980s, Valery Sutton designed Sutton DanceWriting as a system for transcribing dance moves with the help of iconic, deictic and symbolic signs (Sutton 1981/1982). Meanwhile, this way of representing movements of the whole body developed into Sutton SignWriting (SSW) a large and detailed model for transcribing body movements in sign languages around the world, especially the most frequently used manual movements (hand shapes and arm positions, see Figure 1), movements of the torso, spatial movements and facial movements (see Figure 2):

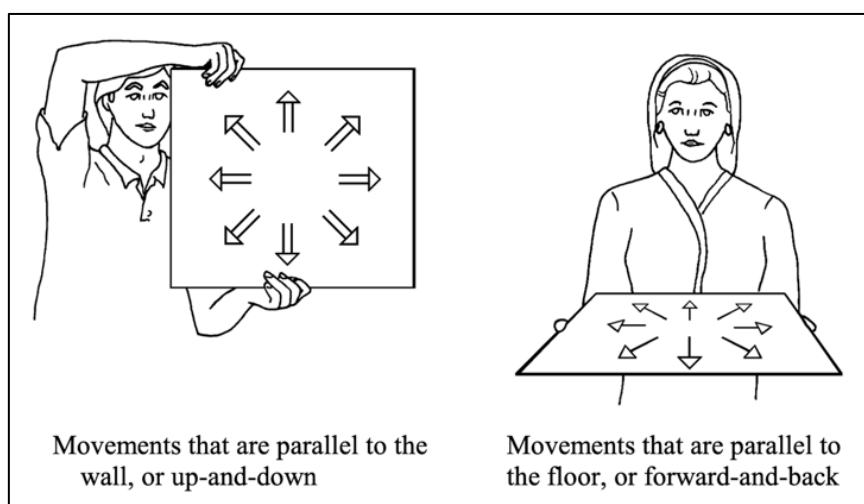**Figure 1.** *ISWA hand signs (Parkurst & Parkhurst, 2010: 29).*
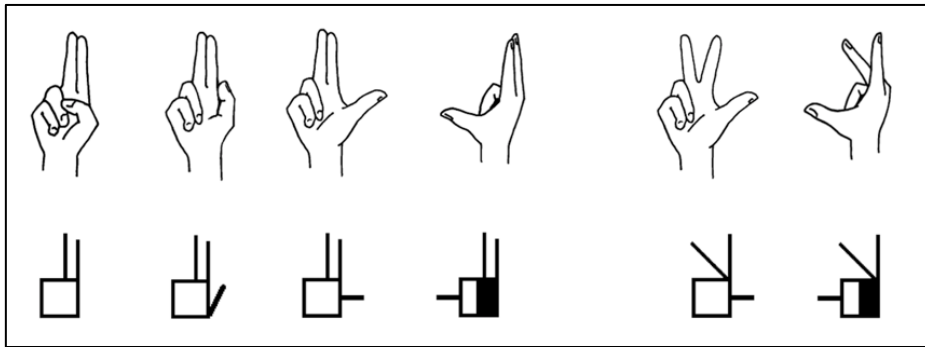
Figure 2. *ISWA signs of facial resources (Sutton, 1982: 82).*
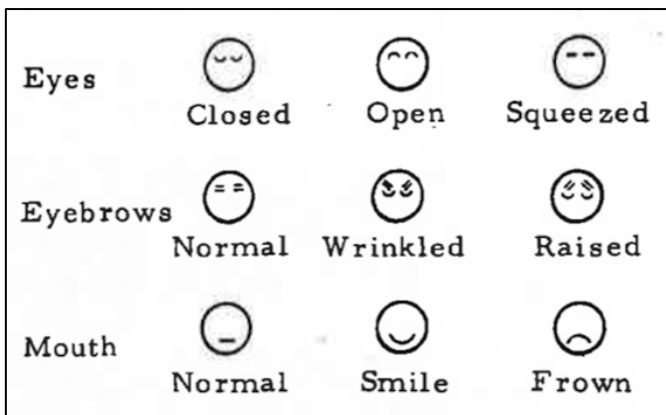
These iconic signs can be depicted as they unfold spatio-temporally by combining them with indexical signs such as arrows and process-oriented signs in order to represent their dynamics. In this respect, arrows, for instance, can cover the following parameters:

- **The direction of the movement**: single shaft arrows ($\hat{|}$) represent movements on the so-called *floor plane* (i.e. forward and backward movement), double shaft arrows ($\hat{\|}$) represent movements on the so-called *wall plane* (up and down movement, see Figure 3)

- **The part of the body that performs the movement**: a black arrowhead ($\hat{\|}$) represents that the right part of the body is moving; white arrowheads ($\hat{\|}$) represent the left part of the body; and open arrowheads ($\hat{|}$) mark that the movement is performed by the whole body, both hands, etc.

- **The form of movement**: straight (←), curved (⌒ɹ), etc.

Figure 3. *ISWA movement signs (Parkhurst & Parkhurst, 2010: 24).*

Movements that are parallel to the wall, or up-and-down

Movements that are parallel to the floor, or forward-and-back

All signs are collected in the International SignWriting Alphabet (ISWA; Parkhurst & Parkhurst, 2010). Although the system provides a huge amount of already standardized signs, the system offers the option to combine different signs for creating composite signs that represent complex or uncommon movements. The ISWA signs are accessible through the platform SignMaker 2022 (created by Steve Slevinski; https://www.sutton-signwriting.io/signmaker/).[3]

Unlike emojis, the iconic, indexical and symbolic signs of the ISWA do not interpret the positions and movements but simply portray their form and the way they are conducted by the interactants in a purely descriptive way. In this respect, the ISWA inventory can be fruitful for different methodological approaches investigating social interaction using written transcripts – beyond sign languages and independent of specific linguistic categories.

Part of the SSW idea is to represent all movements from an expressive viewpoint, that is, from the embodied perspective of the person producing the movement/signing (Parkhurst & Parkhurst, 2010: 2; Hoffmann-Dilloway, 2018). This is seen as 'the default and unmarked visual origo of SW' (Hoffmann-Dilloway, 2018: 90).[4]

---

[3] See the introduction by Steve Slevinski:
https://www.youtube.com/watch?v=TofdVe4xuMs.
[4] Further information and introduction to SignWriting and the ISWA:
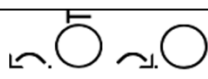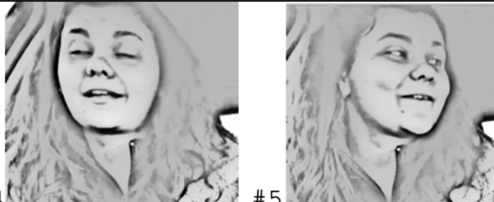https://www.youtube.com/watch?v=ttkMauu_I60 (basic signs, hand shapes, contact symbols)

4. Jefferson Meets ISWA

What I suggest in combining the two elaborate systems of Jefferson and SSW is a rethinking of how conventions which were specifically developed for transcribing visual-bodily behaviour (such as SSW) can be fruitfully used for EMCA and MIA research purposes. The following example illustrates how a transcriber can integrate the Jeffersonian system (2004) with the iconic, indexical and symbolic ISWA signs in a transcript for both analysis and publishing purposes.

The transcript design contains three essential segments (see Figure 4):

- **The verbal transcript:** verbal and vocal elements of the interaction are transcribed according to Jefferson (2004) and ordered in numbered lines according to TCUs. As illustrated in Figure 4, the representation of pauses and silences of a speaker differ from the genuine Jeffersonian conventions, and use dashes to mark a tenth of a second, following the suggestions by Goodwin (1981: vii) and Luff and Heath (2015: 371). The verbal segment can be enriched by an interlinear gloss if grammatical categories of the original language have to be made explicit or to provide an idiomatic translation – see Dix & Groß, 2023/this issue.

- **The visual transcript:** the visual transcript segment consists of several unnumbered lines for representing visual resource levels identified by the researcher as relevant for the interactants within the interaction. The signs are written according to the ISWA and aligned with the verbal transcript and each other by being placed accurately one above the other.

- **Stills** from the original video. Documenting specific prominent and relevant moments. The stills are inserted on unnumbered lines and aligned with the verbal and visual transcript via a number preceded a hash.

Figure 4. *Structure of the transcript used in this article.*

In the example, Kate's (Kte) head movement (line H) and the movements of her eyebrows (line B) are depicted and represented on two separate, unnumbered lines within the visual transcript segment. These lines are placed underneath the verbal segment.[5] This design does not indicate a primary focus on the verbal and vocal aspects of interactions; moreover, it allows the easy extension of already existing verbal transcripts by simply adding the visual lines without modifying the line numbers. Furthermore, lines within the visual transcript can be added or removed – depending on the purpose of the transcript or the focus of the analysis. The actual positions and movements of the resources (here, head and brows) are then indicated by using iconic and indexical ISWA signs.[6] Because the visual actions are perfectly synchronized with the verbal conduct, they are written exactly one above the other.

The last line of the transcript shows two still images from the video. The first presents the moment when Kate's eyebrows reach their highest point (#4), and the second shows the moment when Kate has turned her head to the left (#5). For aligning the stills with the representations of the verbal and visual conduct, the solution developed by Mondada (2019a) to use a hash and a number is adopted. To sustain the readability and clarity of the transcript, the hash is not inserted into the verbal or visual lines but placed on a separate line within the stills section of the transcript. Its placement then indicates the exact position in relation to the verbal and visual signs.

The following examples (Section 5) demonstrate the benefits and validity of the introduced multimodal transcript design by focusing on the opportunities the ISWA inventory offers for representing facial movements (e.g. of the eyebrows,

---

[5] For a more detailed description of this extract see Section 5, Example 1.
[6] See the appendix for an overview of all lines as well as the related ISWA signs used in the transcripts in this article.

the lips and the eyelids) within a multimodal setting, accompanied by verbal, vocal and other visual-bodily resources.

5. Transcribing Facial Gestures

Raising and lowering the eyebrows are the most frequently used facial movements within social interaction (Ekman, 1979). Generally, they are easily observable, but their temporal unfolding can vary between a very quick eyebrow flash lasting only a few milliseconds up to an eyebrow hold lasting several seconds (see Dix & Groß, 2023/this issue). For an analysis of the interactive functions of brow position and brow movement, it is necessary to not only point out that an interactant is raising or lowering the brows but also to represent the temporal unfolding of the movement as well as the orchestration with verbal, vocal and other visual resources. To that end, the system combining Jefferson and the ISWA offers a practical and clear method of representation, as the following example will demonstrate.

Example 1 shows the interaction between two young women – Jenny (Jny) and Kate (Kte) – who are engaged in a car ride, sitting side by side (see Figure 5) and talking about the love interest of a mutual friend.

Figure 5. *Car Ride.*

The visual transcript segment contains lines for the head (line H), the eyebrows (line B) and the lips (line L) of Kate. Dynamic movements are transcribed with arrows to display the direction of the movement and the form of the movement. Furthermore, dots are added to indicate the duration of the movement. The position of the head and the brows are represented with iconic signs (see also the appendix).

Example 1. *'An older guy'*[7]

http://f.kanjo.de/tbMcr

```
74  Kte:   OH:::::;(---)
             oh
```

Kte	H	
	B	
S		#4 #5

```
75       ein ÄLterer.
           an older guy
76       nice.
           nice
```

Kte	B	
	L	
S		#6

After Jenny reports that a mutual friend of the two girls has left her former boyfriend, Kate assesses this friend as being very pretty and concludes that she will certainly find another boyfriend soon (not in the transcript). Jenny then mentions that there already is a new love interest. The transcript shows that Kate verbally reacts by marking the information as new and unexpected with the change of state marker 'oh' (oh:::::, l. 74). Shortly before producing the verbal conduct, she starts raising both eyebrows simultaneously. The movement upwards is indicated by the double shafted arrow (⇑, l. 74/B); the form and position of the brows when they reach the peak position is indicated by an iconic sign (⬭, l. 74/B, #4), representing the stylized head of a person with the position of the brows.

When her eyebrows have risen, Kate starts turning her head towards her left – again indicated by the indexical sign of an arrow; however, the arrow is now single shafted and curved (↰, l. 74/H) – until it is directed to her co-interactant Jenny (⊖, l. 74/H). During this movement, Kate's eyebrows are held in raised position. For a better readability, the transcript represents only changes in position and movement. There is no additional sign used to indicate that the position is held. Rather, it is applicable until another sign marks a change in position. Thus, the raised eyebrows are held until the moment when Kate's head has reached the left-turned position. Then, Kate lowers her eyebrows (⇓), bringing them back to her neutral position. Shortly afterwards, Kate starts moving her head to the right (↱, l. 74/H), bringing it back to neutral position (○, l. 74/H). Kate finally comments on the news by concluding that the love interest must be older (l. 75) and produces a positive assessment ('nice', l. 76) accompanied by an eyebrow flash – that is, a very quick up-down movement of both eyebrows (⇑⇓, l. 76/B) – and an open smile (⌣, l. 76/L). Here, the peak moment is not represented by an ISWA sign but visualized via the still photograph (#6). Therefore, this transcript demonstrates different ways of representing brow positions and brow movements. Depending on the temporal unfolding of the movement – whether it is a continuous move or a hold, how long the raising or lowering lasts and so on – the transcriber can decide on an individual basis whether to use both arrows and icons (l. 74) or only arrows or icons (l. 76).

Beyond the form of the brow positions and the temporal unfolding and movement of the brows, the transcript also depicts the intrapersonal orchestration of the brow movements with other resources such as the spoken word and bodily behaviour, such as the movement of the head and the lips. In this respect, the transcript can lucidly represent that the raising of the eyebrows starts before the verbal production of the 'oh', that the eyebrows are held during the utterance and the head movement, and that the lowering happens in alignment with the end of the verbal production but before turning the head back to neutral position. The second eyebrow movement (eyebrow flash) happens simultaneously with the verbal utterance (l. 76) and an open smile.

Example 2 demonstrates the value of the system when representing a complex multimodal gestalt such as the display of a thinking face (Goodwin & Goodwin, 1986; Heller, 2021), which includes multiple facial resources such as the movement of the lips. Lip movement as an interactive resource in its own right can appear in many different configurations: lips can be protruded, sucked in, moved to the left or right, be bitten by the incisors and so on. In the following example, it is the protrusion of the lips that becomes relevant in the interaction. With the help of the ISWA, these configurations do not need to be described by transcriber comments but can be represented with the help of iconic signs.

The participants in the following extract are three young women (Betty, Anna and Sue – see Figure 6) engaged in a cooking event. Anna participates in the role of a consultant showcasing a food processor which Betty and Sue are going to test in the course of the event.

Figure 6. *Cooking Event.*

At the beginning of the cooking event, the women talk about the habits and preferences each of them has when it comes to preparing food. Betty (Bty) then states that she does not like to think about what to cook every day. While listening, Sue (Sue) produces a thinking face including protruded lips. The transcript shows Sue's visual behaviour at the level of lip movement (line L), brow movement (line B), head position (line H) and eye direction/gaze (line G).

Example 2. *'New recipes'*

http://f.kanjo.de/uPpnQ

```
05    Bty:    also ich mach das nicht gerne.
              That's not something I like to do
06            ich [(.) denk nicht gern drüber na:ch.
              I do not like thinking about that
07    Ana:       [ach so.
                  I see
08    Bty:    ich-(- - - - - - -)
                 I
```

Sue	L	(image) . (image)	
	B	(image)	
	H	(image)	
	G	(image)	
S		#1	

#1

```
09            °hhh° ich nehm gerne rezepte und sag-
              I like taking recipes and say
```

| Sue	L	↓.(image)	
	B	(image)	
	G	(image)	
S		#2	

#2

```
10            oh das klingt lecker das mach ich jetzt.
              Oh that sounds delicious I cook that now
```

After Anna has asked the participants for what they like or dislike regarding the preparation of food (not in the transcript), Betty replies that in general she likes cooking but does not like to look for a new recipe every single day (l. 5–6). Anna responds with the change-of-state token *ach so* ('I see', l. 7; Golato, 2012).

During Betty's turn, Sue starts producing a thinking face, which displays cognitive processing and ongoing recipiency. She tilts her head to the right (⟳, l. 8/H) and starts to protrude her lips – that is, moving both lips forward (☺, l. 8/L). The ISWA inventory offers an iconic sign representing a stylized head and a black dot for the protruded lips. Arrows within the sign mark the direction of the movement (forward) and the body parts involved (both lips simultaneously). Sue holds the protrusion (☺, l. 8/L), which is synchronized with lowering both eyebrows (☹, l. 8/B) at the moment Betty stops talking (#1). Aligned with Betty's inhaling (l. 9), Sue brings her brows back to a neutral position (☺, l. 9/B). Whereas in Example 1 the exact beginning and ending of the process of raising the eyebrows is transcribed in detail, it is only the peak moment of furrowed brows (l. 8/B) and neutral brows (l. 9(B) which are visualized in Example 2. Finally, Betty continues explaining what she likes when cooking while Sue retracts her lips (↓, l. 9/L), bringing them back to a neutral position (☺; l. 9/L). In both cases of protruding and retracting the lips, dots after the arrow mark the duration of the movement. The new position of the lips is represented by an iconic sign. While Sue is holding the protruded position, no other signs are used. Only when the position is modified (l. 9) the change in configuration is indicated by the corresponding ISWA signs. Thus, one can decide to either simply mark the position of a facial resource or one can more accurately represent the starting point, the stroke and the moment of retraction back to a neutral position.

Whereas the moment of protruding the lips is displayed by a sign version where the arrow and the iconic lip display are included in one symbol, now the arrow and the final lip position are separated for marking the start of the movement more clearly. In this respect, the transcript again demonstrates how signs and sign combinations can be used to portray different facial resources and different degrees of detail, depending on the research question and the purpose of the transcript.

While Sue moves her lips, her gaze turns from an up-left position (☺, l. 8/G) to a straight-right position (☺; l. 9/G). With the help of the ISWA inventory, changes in gaze direction can be transcribed using a sign showing a circle representing a stylized head and arrows representing the direction of movement (up or down) or the gaze direction (straight, left, right). Here, the arrows are placed within the circle to clearly mark that it is the eyes moving.

The transcript shows that the thinking face is a complex configuration of different facial resources (eyebrows, lips) and other resources of the upper body (gaze,

head). By depicting the individual resources which are involved to produce the thinking face, one can detail the interplay of brow position/movement, lip position/movement, gaze and head position, and represent how these resources are aligned with each other as well as with the spoken conduct of the producer or the recipient of the facial gesture.

Whereas the examples discussed so far have demonstrated how to represent the intrapersonal coordination of different verbal and visual resources, Example 3 illustrates how to depict the interpersonal coordination of visual resources. Matthew (Mtw) and Lisa (Lsa) are engaged in a dinner talk, which was recorded with a 360° camera. The transcript traces the levels of the interlocutors' heads (line H), their torsos (line T), their eyebrows (line B) and their eyelids (line E).

Example 3. *'Lady and the Tramp'*

http://f.kanjo.de/UkbSw


```
42  Mtw: susi und strolch?(---)
             Lady and the tramp?
```



```
43  Lsa: wie bitte?
             Pardon?
```



```
45  Mtw: in deiner jugend,(----)
             In your youth
```

While they are having dinner, Matthew scrolls through a cookbook containing recipes from famous animated films, for instance, meat balls with spaghetti from the film *Lady and the Tramp*. Being unsure about the release date of the movie, he turns to Lisa, moving his upper body orientation to up-left (⟶, l. 42/T-Mtw) and his head direction up-left too (◯, l. 42/H-Mtw) towards Lisa until his head is in a left oriented position (◯, l. 42/H-Mtw). Simultaneously, he names the title of the film with raising intonation, marking it as a first part of a question-answer-sequence ('lady and the tramp', l. 42). Lisa turns her head to the right (◯, l. 42/H-Lsa), which is followed by leaning her torso to the right (⟶, l. 42/T-Lsa), bringing it closer to Matthew and with that establishing a focused interaction with him. Shortly after she has started moving, she verbally initiates repair ('pardon', l. 43). Her head is now oriented towards Matthew (◯, l. 43/H-Lsa) and her torso leans to the right (⟶, l. 43/T-Lsa). Instead of following the repair initiation, Matthew hints at the life phase when Lisa could have seen the film ('in your youth', l. 45). Lisa again initiates repair by raising her eyebrows (⇑, l. 45/B-Lsa) and holding them raised (◯, l. 46-B-Lsa) until Matthew has repeated the title of the film ('lady and the tramp', l. 46) and his question about the release date ('when was it', l. 48).

The example demonstrates the fine tuned intrapersonal and interpersonal orchestration of verbal and several visual-bodily resources for the purpose of establishing a focused interaction by marking recipiency (moving the [upper] body and shifting head orientation) and initiating repair (by leaning forward and raising the eyebrows).[8]

6. Benefits and Limitations of Combining Jefferson with ISWA

The examples have demonstrated that by using the iconic, deictic and symbolic signs of the ISWA inventory and combining them with Jeffersonian conventions, the multimodal configuration and orchestration of social interactions can be presented in a clear, (internationally) readable and at the same time precise and detailed way. Hence, the system presented here covers numerous requirements of EMCA multimodal transcripts (see Section 2):

- Firstly, the system is adaptable to different data sets and settings (here, a car ride, a cooking event and a multiparty dinner conversation) as well as different research interests – be it the investigation of functions of a specific single visual resource (e.g. raising both eyebrows) or the

---

[8] See also Pfeiffer & Stolle, 2023/this issue.

complex multimodal gestalt of a gesture (e.g. the thinking face) within its multimodal embeddedness.

- Furthermore, the transcription design can be adapted to different transcription purposes (working transcript or presentation transcript) by adding or deleting lines on the level of verbal transcription as well as on the level of visual transcription. The design can also be adapted to different transcription formats (line transcript, score transcript) as well as different conventions for transcribing verbal and vocal conduct (e.g. Jefferson, GAT2, etc.).

- Thirdly, the ISWA inventory opens the way for representing different kinds of visual-bodily resources (manual as well as non-manual gestures). Even subtle phenomena such as the eyebrow flash can be portrayed in a detailed manner, including their temporal and sequential unfolding as well as their coordination with verbal conduct and other visual-bodily resources. By applying the combination of the Jefferson system and ISWA to the depiction of facial gestures within multimodal transcripts, one can represent the form of the resources, the dynamics of their movements as well as the fine-grained orchestration with other resources.

- By expanding the signs with simple symbols like dots for marking the duration of the movement, synchronizing them with speech and other visual conduct as well as still images, the written representation of the analysis can be refined. The ISWA inventory allows for representing the visual behaviour of the interactants within the transcript – and beyond that – within the analytic description of what happens in the extract. This enables a strong link between the transcript and the interpretation of the data.

- Perhaps the most striking benefit of the presented way of doing multimodal transcription is that the ISWA signs allow a detailed representation of visual conduct without using transcriber comments. Because the ISWA system tries to maintain the iconicity of the visual phenomenon, the signs are not only intuitively readable and easy to learn, but consequently circumvent metacommunicative transcriber comments. As a result, the transcripts shown above are all free of verbal descriptions or interpretations of the movements and positions used by the interactants.

On a technical level, the signs of the ISWA inventory can easily been accessed via SignMaker 2017[9] or SignMaker 2022.[10] These programs allow the extraction and downloading of the signs as PNG or SVG images (see the video via the QR code for more detail on how to extract the images. Link: http://f.kanjo.de/IxZxk). Furthermore, the programs allow the combination of several signs.

However, one of the challenges is that unfortunately the image files of the ISWA signs currently cannot be integrated into transcription software such as Elan.[11] Therefore, for the purpose of this contribution, the signs have been accessed via SignMaker 2022, then extracted and downloaded so that it was possible to integrate them into the transcripts in Word.

Another challenge is that the ISWA inventory, as detailed as it may be, has been optimized for the purpose of representing sign languages. Therefore, it offers a detailed way of transcribing manual and non-manual gestures, the position and movements of the whole body as well as its orientation and intrapersonal contact movements – such as touching one's own face, hands, arms and so on. At the same time, this also means that handling objects and interpersonal touch are not easy to represent. To that end, further ideas and creative ways of expanding the ISWA inventory are needed.

7. Conclusion

The aim of this contribution was to provide a practical guide for transcribing visual-bodily behavior aligned with verbal and vocal conduct by combining Jeffersonian conventions and the ISWA system. In this respect, it was shown that the iconic ISWA signs are flexible, readable and adaptable enough to be used for transcribing naturally occurring social interaction. By giving the example of transcribing facial gestures, it was shown how the comprehensiveness and exhaustiveness of the ISWA sign repertoire could become part of a standardized and at the same time research-oriented convention of multimodal transcription. The proposed system aims at finding a balance between the transcriber's

---

[9] http://www.signbank.org/signmaker/#?ui=de&dictionary=gsg [latest access: 26.03.2023]

[10] https://www.sutton-signwriting.io/signmaker/ [latest access: 26.03.2023]

[11] The technical support team of Elan is aware of the need to integrate images into Elan and is working on a solution.

research focus and the complexity of the data, with the need for highly dynamic transcripts on the one hand and the need for a standardized system making data accessible to other researchers and helping them to read, comprehend and compare data and research results on the other hand. Although there are already numerous benefits to the system, working on the challenges that have appeared will further improve it.

References

Aaarsand, P. & Sparrman, A. (2021). Visual transcriptions as social-technical assemblages. *Visual Communication. 20*(2). 289-309.

Albert, S. et al. (2019). Drawing as transcription: How do graphical techniques inform interaction analysis? *Social Interaction. Video-Based Studies oh Human Sociality.* 2(1). https://tidsskrift.dk/socialinteraction/article/view/113145/161800 (21.04.2022)

Ayaß, R. (2015). Doing data: The status of transcripts in Conversation Analysis. *Discourse Studies. 17*(5). 505-528.

Bezemer, J. (2014). Multimodal transcription. A case study. In: Norris, S. et al. (Eds.), *Interactions, Images, Texts. A reader in multimodality*. Berlin, Boston: deGruyter. 155-169.

Bezemer, J. & Mavers, D. (2011). Multimodal transcription as academic practice. A social semiotic perspective. *International Journal of Social Research Methodology*. 14(3). 191-206.

Birdwhistell, R. L. (1973). *Kinesics and Context: Essays on body-motion communication*. Middlesex: Penguin.

Bohle, U. (2013). Approaching notation, coding, and analysis from a conversational analysis point of view. In: Müller, C. et al. (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (HSK 38.1). Berlin u.a: deGruyter. 992-1007.

De Stefani, E. (2022). On Gestalts and their Analytical Corollaries: A Commentary to the Special Issue. *Social Interaction. Video-Based Studies of Human Sociality.* 5(1). Online: https://tidsskrift.dk/socialinteraction/article/view/130875/177121 (10.05.2022).

Dix, C. (2021). *Die christliche Predigt im 21. Jahrhundert: Multimodale Analyse einer Kommunikativen Gattung*. Wiesbaden: Springer VS.

Dix, C. (2022). GAT 2 trifft das International SignWriting Alphabet (ISWA): Ein neues System für die Transkription von Multimodalität. In: Schwarze, C. & Grawunder, S. (Eds.), *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion: Konzepte, Probleme, Lösungen*. Tübingen: Narr. 103-131.

Drew, P. (2017). Gail Jefferson and the development of transcription. In: Hepburn, A. & Bolden, G. (Eds.), *Transcribing for social research*. Los Angeles: Sage. 5-6.

Ekman, P. (1979). About brows. Emotional and conversational signals. In: Cranach, M. von, et al. (Eds.), *Human ethology. Claims and limits of a new discipline*. Cambridge University Press, 169-202.

Ekman, P. & Friesen, W. & Hager, J. C. (2002). *Facial Action Coding System: An ebook for pdf readers*. Douglas/Arizona: a Human Face.

Golato, A. (2012). German oh: Marking an Emotional Change of State. *Research on Language and Social Interaction*. 45(3). 245-268.

Goodwin, C. (1981). *Conversational Organization. Interaction between speakers and hearers*. New York (u.a.): Academic Press.

Goodwin, C. (2000): Practices of Seeing. Visual Analysis: An Ethnomethodological Approach. In: Leeuwen, T. van & Jewitt, C. (Eds.), *Handbook of Visual Analysis*. London: Sage. 157-182.

Goodwin, M. H. & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*. 62(1-2). 51–75.

Heath, C. et al. (2010): *Video in Qualitative Research. Analysing social interaction in everyday life*. Los Angeles: Sage.

Heller, V. (2021). Embodied displays of 'doing thinking'. Epistemic and interactive functions of thinking displays in children's argumentative activities. *Frontiers in Psychology*. 12. 1-21.

Hepburn, A. & Bolden, G. (2017). *Transcribing for social research*. Los Angeles (u.a.): Sage.

Hoffmann-Dilloway, E. (2018). Feeling your own (or someone else's) face: Writing signs from the expressive viewpoint. *Language and Communication*. 61. 88-101.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In: Lerner, G. H. (Ed.): *Conversation Analysis. Studies from the first generation*. Amsterdam, Philadelphia: John Benjamins. 13-31.

Jenkins, C. (2018). Recording and transcribing social interaction. In: Flick, U. (Ed.): *The SAGE handbook of qualitative data collection*. Sage. 118-130.

Kendon, A. (2004). *Gestures. Visible Action as utterances*.

Kennedy, A. (2013). Laban based analysis and notation of body movement. In: Müller, C. et al. (Eds.), *Body – Language – Communication. An international Handbook on Multimodality and Human Interaction*. (HSK 38.1). 941-958.

Knoblauch, H. & Heath, C. (2006). Die Workplace Studies. In: Rammert, W. & Schubert, C. (Eds.), *Technografie. Zur Mikrosoziologie der Technik*. Frankfurt, New York: Campus. 141-161.

Laurier, E. (2014). The Graphic Transcript: Poaching comic book grammar for inscribing the visual, spatial and temporal aspects of action. *Geography Compass*. 8(4). 235-248.

Luff, P. & Heath, C. (2015). Transcribing embodied action. In: Tannen, D. et al. (Eds.): *The Handbook of Discourse Analysis*. John Wiley. 367-390.

Mittelberg, I. (2007). Methodology for multimodality. One way of working with speech and gesture data. In: Gonzalez-Marquez, M. et al. (Eds..), *Methods in Cognitive Linguistics.* Amsterdam. Philadelphia: John Benjamins. 225-248.

Mondada, L. (2007). Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse Studies*. 9(6). 809-821.

Mondada, L. (2017). Multimodal transcription and the challenges of representing time. In: Hepburn, A. & Bolden, G. (Eds.), *Transcribing for social research*. Los Angeles: Sage.

Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction. Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*. 51(1). 85-106.

Mondada, L. (2019a). *Conventions for multimodal transcription*. Online: https://www.lorenzamondada.net/multimodal-transcription (10.05.2022).

Mondada, L. (2019b). Transcribing silent actions: A multimodal approach of sequence organization. *Social Interaction. Video-Based Studies of Human Sociality.* 2(1). Online: https://tidsskrift.dk/socialinteraction/article/view/113150 (10.05.2022)

Parkhurst, S. & Parkhurst D. (2010). *A cross-linguistic guide to SignWriting. A phonetic approach*. Online:

http://www.signwriting.org/archive/docs7/sw0617_Cross_Linguistic_Guide_SignWriting_Parkhurst.pdf (21.04.2022).

Rossano, F. (2013). Gaze in Conversation. In: Sidnell, J. & Stivers, T. (Eds.), *The Handbook of Conversation Analysis*. Oxford: Blackwell. 308-329.

Sager, S. (2001). Probleme der Transkription nonverbalen Verhalten. In: Brinker, K. et al. (Eds.), *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. (HSK 16.2). 1069-1085.

Selting, M. et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353-390. Retrievable online: http://www.gespraechsforschung-online.de/heft2009/heft2009.html (retrieved 30th Oct. 2020).

Streeck, J. (2009). *Gesturecraft: The manufacture of meaning.* Amsterdam, Philadelphia: John Benjamins.

Stukenbrock, A. (2009). Herausforderungen der multimodalen Transkription: Methodische und theoretische Überlegungen aus der wissenschaftlichen Praxis. In: Birkner, K. & Stukenbrock, A. (Eds), *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung*. Mannheim: Verlag für Gesprächsforschung. 144-169.Online abrufbar: http://www.verlag-gespraechsforschung.de/2009/birkner.html (30.10.2020).

Sutton, V. (2010). *The SignWriting Alphabet. Read and Write any Sign Language in the World. ISWA Manual 2010*. The SignWriting Press. http://www.movementwriting.org/symbolbank/ (30.10.2020).

Sutton, V. (1981/1982). Sutton Movement Writing and Shorthand. *Dance Research Journal*. 14(1 & 2). 78-85.

Wagner, J. (2020). Conversation Analysis: Transcriptions and Data. In: Chapelle, C. (Ed.), *The concise encyclopedia of applied linguistics*. Wiley-Blackwell. 296-303.

Waller, B. M. & Smith Pasqualini, M. (2013). Analysing facial expression using the facial action coding system (FACS). In: Müller, C. et al. (Eds.): *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (HSK 38.1). Berlin u.a: deGruyter. 917-931.

Appendix – Transcription

B   Eyebrows

E   Eyelids

G   Eyes / Gaze

H   Head

L   Lips

T   Torso

ISWA Signs used in the transcripts:

Head straight

Head tuned to the right

Head turned to the left

Head turned right-up

Head leaned to the right

Head moves to the right

Head moves up-left

Brows straight

◠ Brows raised

◡ Brows lowered

⊖ Lips straight

☺ Open smile

Lips moving forward

Lips protruded

Eyes directed up-left (wall plane)

Eyes directed straight-right (floor plane)

➜ Torso moves to the right

➘ Torso moves up-left

➚ Torso leans right-forward

↶ Movement to the left

↷ Movement to the right

⇑ Movement upward (whole body, both arms, both eyebrows etc.)

⇓   Movement downward (whole body, both arms, both eyebrows etc.)

⇑   Movement upward (left body part, left arm, left eyebrow etc.)

⇑   Movement upward (right body part, right arm, right eyebrow etc.)

↑   Movement forward

↓   Movement backward

←   straight Movement to the left (right body part, right arm etc.)