Social Interaction. Video-Based Studies of Human Sociality.
2021 Vol. 4, Issue 3
ISBN: 2446-3620
DOI: 10.7146/si.v4i3.128151

Social Interaction

Video-Based Studies of Human Sociality

Making sounds visible in speech-language therapy for aphasia

Sara Merlino

Università degli Studi Roma Tre

Abstract

In this paper, I analyse video recordings of speech-language therapy sessions for people diagnosed with aphasia. I particularly explore the way in which the speech-language therapists instruct the patients to correctly pronounce speech sounds (e.g. phonemes, syllables) by deploying not only audible but also visible forms of cues. By using their bodies – face and gestures – as an instructional tool, the therapists make visual perceptual access to articulatory features of pronunciation relevant and salient. They can also make these sensory practices accountable through the use of other senses, such as touch. Data was collected in a hospital and in a rehabilitation clinic, tracking each patient's recovery, and is part of a longitudinal multisite corpus. The paper considers the way in which participants in the therapeutic process use and coordinate forms of sensory access to language that are based on hearing and seeing. It highlights the importance of audio and video recordings to make accessible the auditory and visual details of these sensorial experiences – particularly, proper framings and the complementary use of fixed and mobile cameras.

Keywords: speech-language therapy, aphasia, pronunciation, instructed vision, auditory and visual resources, multimodality, sensoriality, video-camera framings

1. Introduction

A variety of medical and therapeutic settings have been investigated with the use of ethnomethodological and conversation analytic video-based research (among many[1], for example, for doctor-patient interactions, see Heath, 1986, 2018; for prenatal ultrasound examinations, see Nishizaka, 2011; and for physiotherapy sessions, see Parry, 2005). The therapeutic setting of speech-language therapy (SLT) for adults with aphasia has also benefited from the use of video recording for documenting and analysing the role of resources other than speech in the therapeutic process. In particular, emphasis has been put on the aphasic person's use of embodied resources (such as gestures, gaze or facial expressions) during the therapy, mainly in relation to specific activities or actions, such as word searches or topic initiations (Klippi, 2015; Klippi & Ahopalo, 2008; Laakso & Klippi, 2010; Merlino, 2017; Wilkinson, 2011). Moreover, the implementation of these resources by the patient has been advocated in conversation therapy programmes (see Beeke et al., 2015). In this paper, I broaden this line of investigation by considering the way in which the *therapist* uses embodied resources during the session, particularly in order to instruct the patient to correctly pronounce specific speech sounds (phonemes or syllables) as part of language recovery. During these moments, the pronunciation of linguistic items is treated and experienced not only as an audible but also as a visible phenomenon: indeed, the therapist uses his/her body as an instructional tool by proposing to the patient forms of sensorial access to language that are based on hearing and seeing and that are conveyed through verbal, gestural and, sometimes, haptic (e.g. touch) means.

Research on multimodality has largely explored the visual dimension of communication and described practices of seeing. Thanks to the use of video recordings of natural occurring interactions taking place in their material environment, a praxeological and intersubjective approach to visual perception and visual skills has developed (Goodwin, 1981, 1994; Kidwell, 2005; Mondada, 2014; Rossano, 2012; Streeck et al., 2011). This approach treats the practices of seeing in relation to a cluster of resources, such as participants' bodies and embodied postures, the use of artefacts and tools, the establishment of participation frameworks (see Mondada, 2018a, for a discussion). The practices of hearing and the auditory dimension of communication have been investigated in terms of audibility of the linguistic units (see the large literature on repair, cf.

[1] See Pilnick et al. (2009) for a review of CA studies that explore medical settings and healthcare interactions.

Jefferson, 2017; on the topicalisation of hearing loss, see Egbert & Deppermann, 2012). More recently, an interest in the perception of non-linguistic sounds has developed, with analyses that take into consideration how participants orient to sounds of the body (for sniffs or grunts, see Keevallik & Ogden, 2020) or of the environment (Merlino et al., forth.). This reflects a recent interest in interactional studies in sensory dimensions such as taste (Liberman, 2018) and tactility (Cekaite & Mondada, 2020; M.H. Goodwin, 2017; M.H. Goodwin & Cekaite, 2018) and an expansion of the use of multimodal analysis into the realm of multi-sensoriality (Mondada, 2019, forth.).

The relationship between the visual and auditory dimensions of communication and the senses that realise them (hearing and sight) are especially interesting in the treatment of speech and language impairments, such as aphasia. My concern is the interactive work realised by the participants, particularly by the therapist, in order to experience, share and coordinate different forms of sensorial access to language in the accomplishment of the therapeutic process (for other forms of sensorial access and 'embodied practices for sensing the world in an intersubjective way', see Mondada, 2019, p. 47). As the analysis of data will show, the possibility of analysing these forms of sensorial access is connected to what is made available by the audio and video recordings collected during the fieldwork. I will particularly point to the technical choices made with respect to camera angles and positioning of a separate audio recorder in order to capture the participants' deployments of audible and visual resources.

2. Analysing speech and language therapy through audio and video recordings

The documentation of aphasia speech-language therapy through audio and video recordings has allowed for a large body of work in the field of pragmatic and interactional research on the communicative dynamics of this setting. Among other things, the interactive (see, for instance, Klippi, 2015; Laakso, 2015; Simmons-Mackie & Damico, 1999; Wilkinson, 2004, 2011) and the institutional and instructional dimensions of the therapy have been highlighted (Ferguson & Armstrong, 2004; Horton & Byng, 2000). Although the analyses have mainly focused on the verbal exchanges of participants, some researchers have also recognised and explored the role of embodied and multimodal resources in the accomplishment of the therapeutic process (Klippi, 2015; Klippi & Ahopalo, 2008; Laakso & Klippi, 2010; Merlino, 2017; Wilkinson, 2011).

Interestingly, in aphasia literature, there are important claims about the role and efficacy of multimodality in the treatment of patients' linguistic and communicative skills (Dunn, 2010; Pierce et al., 2019; Rose & Attard, 2011). However, these results have generally not included the detailed, frame-by-frame analyses of video recordings – although these results can be based on videos (indeed, the collection of video recordings does not necessarily imply a detailed analysis of

participants' transcribed conduct). Moreover, the focus of aphasia literature regarding multimodality is not, in general, on the turn-by-turn *interaction* between therapists and patients but rather on the single multimodal 'performances' of each participant – the therapist or the patient (see the useful distinction by Peirce et al. (2019) between 'input, therapist cueing and patient output').

Yet the detailed analysis and multimodal transcription of video recordings allow the showing not only of how participants' joint action, mutual understanding and organisation of the therapeutic activities are performed through a cluster of audible and visible resources; it also permits analysis of how different senses are involved and coordinated in the therapeutic experience. In the hospital setting (particularly, in stroke units), I noticed, for instance, that speech-language therapists used haptic resources, such as (self and other) touch, to instruct the patient about appropriate articulatory movements and, during the accomplishment of the linguistic tasks, to manage his/her attention or delicate moments such as manifestations of frustration, anger or crying (Merlino, 2020, forth.). The (professional) use of the therapist's body as an *instructional tool* is indeed an essential part of his/her work: its analysis can shed light on how 'multimodal therapies' are concretely implemented by participants. In this paper, I describe how the therapist makes visible, highlights and emphasises the articulatory movements for the pronunciation of the linguistic items to be produced by the patient. Embodied and visual resources are then not only used for communicating, but also to instruct how to 'feel/perceive' language by displaying, in a therapeutic setting, a sensory access to language and its units that is based both on hearing and seeing.

3. Making sounds visible and instructing vision in institutional settings

There are two interrelated aspects concerning the use of the body by the therapists to instruct the patients about features of linguistic units through the visual modality: first, the way in which the therapists manage to direct the patients' attention and gaze towards their body and face in order to establish a certain perception of the visual cues; second, the specific embodied resources (such as facial and hand gestures) the therapists use to visually represent the linguistic items.

Visual practices have been largely investigated in video-based studies of social interaction, in which methods of seeing, looking and gazing have been described in different types of contexts (e.g., Goodwin, 1981; Kidwell, 2005; Mondada, 2014, 2018a; Rossano, 2012). Particular attention has been paid to the way in which a common visual perception and joint attention are established by participants through interaction using both verbal (Kidwell & Zimmerman, 2007) and embodied resources (Goodwin, 2003; Mondada, 2014). Among these latter resources, pointing gestures play a crucial role: together with language, gaze and

body, they allow participants to not only negotiate reference but also to instruct about relevant features of the surrounding environment for understanding and accomplishing the task at hand (Goodwin, 1994). In instructional activities in which the object of instruction is a bodily practice, recipient's gaze can be treated as crucial in order to pursue the activity and deliver an embodied instruction (Svensson et al., 2009) or for an 'instructed perception' (Nishizaka, 2014). Here, 'the request for gaze assumes a prominent function. As the participants orient themselves to their co-participants' bodily demonstrations, perception and reciprocity of perception play a constitutive role for the organization of a shared perceptual and embodied common ground' (Stukenbrock, 2014, p. 97).

As far as instructions about speech 'performances' are concerned, experimental research has highlighted the role of gestures in the L2 acquisition process (for a synthesis, see Gullberg, 2006). Even if it is mainly the relation among gestures and acquisition of the lexicon that have been addressed (Lazaraton, 2004), the effect of gestures and lip movements on L2 learner comprehension has also been explored (Sueyoshi & Hardison, 2005). By looking more attentively at the communicative dynamics of classroom interactions, Smotrova (2017) highlighted the role of the body as a pedagogical tool in teaching pronunciation of a second language. The author recognised the central role of teachers' gestures and body movements for facilitating students' identification and production of suprasegmental features, such as stress and rhythm. Gestures allow one to visualise and experience pronunciation phenomena: according to Smotrova (2017), students respond to gestures by repeating and mirroring them while at the same time trying to reproduce the target pronunciation features.

As far as the relation between visual cues and word retrieval in aphasia therapy is concerned, experimental research has suggested the efficacy not only of therapists' gestures (see, for instance, Rose, 2013), but also of visual cues such as lip position (what is technically called 'visual phonemic cueing'). On the basis of experiences in which the aphasic speaker is assisted by a computer that provides for audible and visible cues such as mouth shape, Fridriksson et al. (2009) showed, for instance, that audio-visual treatments worked better than audio-only treatments. My study highlights the connection between the auditory and visual perception of speech in the course of the therapeutic activity. It points to the praxeological and intersubjective way in which participants, in order to achieve their therapeutic purposes, experience and share perception of speech sounds through auditory and visual modalities.

4. Choosing relevant camera angles for capturing the auditory and visual details of interaction

Research grounded in video-based methodology invites the consideration of the deep interconnection between the way data is collected and filmed and the type of analysis that can be made: that is, the analytical implications of the technical

choices made during the recordings (Goodwin, 1981, 1994; Mondada, 2006). Despite technological developments and methodological reflection on video data collection (see Heath, Hindmarsh & Luff, 2010), selecting a proper angle and framing remains a crucial and still debatable issue: 'To obtain data that are amenable to analysis, the importance of what might seem mundane choices, such as at what height to place the camera, where to point the lens and how wide to set the focus, can be critical for the subsequent analysis that can take place' (Luff & Heath, 2012, p. 273). These technical choices are contextual by definition and are thus dependent on the constraints of the scene, the participants' dispositions, the time at the disposal of the researcher for arranging the camera setting and the researcher's knowledge of the scene and activity (see also LaBonte et al, 2021/this issue). The researcher needs to capture not only *that* participants collaborated in an activity but also *how* they did so in order to open participants' perceptions of the environment to at least partial scrutiny (Luff & Heath, 2012). To this end, the use of a mid-shot camera, along with multiple other cameras, is an effective choice – even if the specificities and complexities of each setting to be recorded always require local adjustment and reflection (see the discussion by Luff & Heath, 2012).

The data presented in this paper were collected in two different settings – a hospital and a rehabilitation centre – and are part of a large corpus of longitudinal data (approximately 60 hours) that I collected during the six-month recovery period of several aphasic patients who had suffered a stroke. The corpus allowed the documentation of speech and language therapy sessions that took place in different contexts (hospital, rehabilitation centre, private office, patient's home), at different moments of the patient's recovery and with different speech and language therapists. In total, thirteen patients and nine speech-language therapists were filmed. This multisite collection of data allows the recognition, under the same heading ('speech-language therapy'), not only of a variety of therapeutic objectives (e.g. early recovery vs routine therapy for chronic disease) but also a diversity of material and contextual features of the different settings. For instance, I observed variety in the way participants were positioned (e.g. sitting vs standing, one in front of the other vs side by side) and in the type of artefacts they used (such as objects, cards, documents or a computer). The features of these ecologies of action have an impact on participants' body conduct, visual orientation and, finally, on the organisation of the activity itself (Goodwin, 2000).

At the hospital and in the rehabilitation clinic, the speech-language therapy sessions took place in a small room. Two participants – the therapist and the patient – or a maximum of three participants[2] were present. In both of these cases, I decided to use two cameras. One was a fixed mid-shot camera positioned on a tripod: according to the configuration of the room, this camera

[2] When an apprentice was present (like the woman in extract 2, see section 5.3).

sometimes was positioned slightly above the participants' heads, angled down in order to capture the activities of both participants from 'one side' of the scene. Figures 1 and 2 show the use of this camera in the two settings:

Figure 1. Hospital room – fixed camera

Figure 2. Rehabilitation clinic – fixed camera

The other camera was a mobile camera that I held myself, as I was present during the filming and was following the session, either sitting or standing in the room. The camera was generally positioned at the level of participants' upper bodies (lower than their heads), offering a perspective of what the participants had access to and what they could see. Even if this camera was generally kept 'static', with a view of both participants, it benefited from my online analysis of what participants were doing (and so, eventually, I could adjust to changes in the activity, zooming quickly on artefacts they used or following interruptions due to

someone entering the room, which was quite frequent in the hospital setting). Figures 3 and 4 show the vision offered by this camera in the two settings:

Figure 3. Hospital room – mobile camera

Figure 4. Rehabilitation clinic – mobile camera

This second camera allowed the capture of the scene in a *complementary* way with respect to the other camera in order to get the best possible access to participants' mutual actions and perceptions of the environment (that is, what they had access to). Thanks to my proximity to participants (with good access to participant activities), I could capture both the participants' upper bodies and faces and focus, in particular, on the participants whose faces were not clearly visible in the other camera view. Each camera then made it possible to document, in a complementary way, the details of each participant's visual conduct.

The sound was recorded with a separate voice recorder in order to obtain a better audio quality than the one offered by the camera. The recorder was positioned on the table used by participants, as closely as possible to them, to capture the participants' audible productions with minimal interference of sounds and noises from the environment.

This access to participants' faces and upper bodies as well as a good sound recording allowed me to capture visible and audible details that, transcribed in detail through multimodal transcriptions (Mondada, 2018b), allow the type of analysis I present in this paper. A discussion could be opened about the (irresolvable?) discrepancy between what is *de facto* perceivable and perceived by participants during the accomplishment of their actions and what is captured by the cameras and audio recorders. However, a camera set that properly captures participants' vocal, verbal, facial and gestural productions allows an analysis and discussion of the ways in which different senses are coordinated in the therapeutic process and paves the way for analyses of sensoriality as publicly and intersubjectively organised (see, on touching, Cekaite & Mondada, 2020; Mondada et al., 2021/this issue).

The data have been transcribed following the conventions developed by the ICAR Research Lab for the verbal data and by Mondada (2018b) for the visual data (see references in the appendix).

5. Analyses

In what follows, I analyse some sequences of different therapeutic activities (words/syllable repetition, picture-naming and reading aloud) that were aimed at the production of speech. I first analyse the way the therapist, through *pointing gestures*, directs the patient's attention to the therapist's face (the mouth) and makes relevant specific visual details for the accomplishment of the task and the pronunciation of the target item (section 5.1). I then focus on the *embodied resources* (representative gestures and mouth expressions) that are used to represent the target sounds and make them recognisable and reproducible (section 5.2). Finally, I show that the therapist, while (s)he is offering visual support to the patient, can handle, also through forms of touch, concurrent foci of visual attention, which are due to the contextual configuration of the setting and of the task (section 5.3).

5.1. From speech sounds to vision: making pronunciation visible

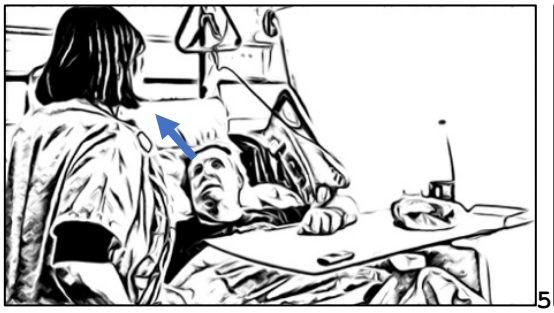
A recurrent pattern observed in my data is the use of (self) pointing gestures by the therapist in order to focus the attention of the patient on his/her face. This is generally done when the patient manifests specific difficulties in the accomplishment of the task, particularly the incapacity of pronouncing one or

more phonemes. When the use of *audible cues* by the therapist is ineffective in helping the patient to correctly pronounce the target sounds, and after several attempts to do so, the therapist opts for a visual representation of the pronunciation. This works both as a form of correction and a new type of hint.

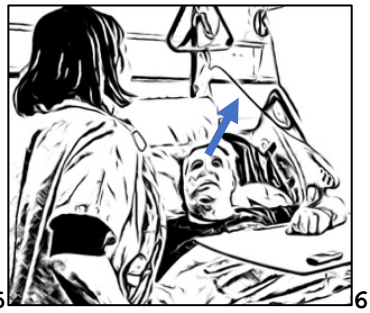
In the first extract I analyse, the participants are involved in an oral activity that consists of repeating a series of syllables: the therapist utters the target syllable, and the patient responds by repeating it. The therapist is standing close to the patient, who is lying in his bed (the session is taking place at the hospital). The arrangement of the participants and the type of activity favour a mutual gaze orientation (Fig. 5). This orientation is nevertheless soon modified by the patient. Following his difficulties in repeating the syllable proposed by the therapist ('BA'), the patient finally modifies the direction of his gaze (line 3):

Extract 1a.

```
1  SLT    .h: (0.2) BA,
   slt    >>lks pat-->
   pat    >>lks slt-->
2         (.)
3  PAT=>  #.h: (0.3)+#(0.3) euh:: euh +haH:
                  --->+gazes up-------->+gaze down-->
   fig    #fig5       #fig6
```

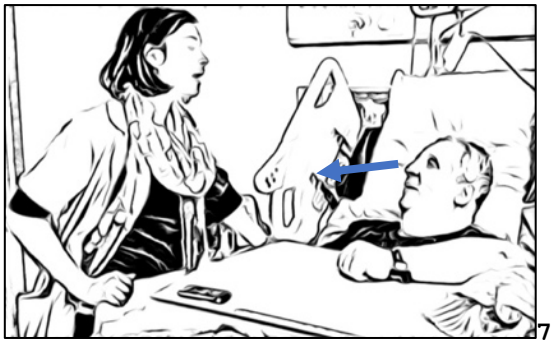


5



6

```
4         +(0.2)
   pat    +turns head,gazes in front-->
5  SLT    #BA.
   fig    #fig7
```



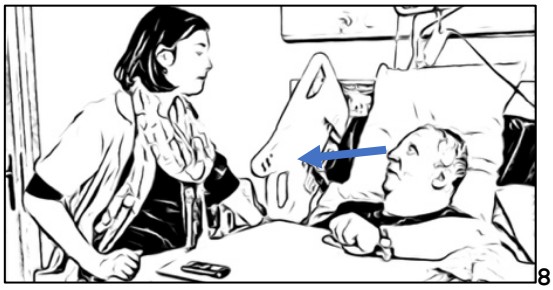
7

```
6         (0.4)
7  PAT    +.h:: (0.4) euh:+>ka ka< k-
   pat    +turns h, glances tw slt+gazes in front-->
8  SLT    <((hold with closed lips,without release of air))b:> BA
```

9 Ω#(0.6)

slt Ωlabial (b)position-->l.15

fig #fig8



8

10 PAT mh mh ka+:

--->+...

The turn at line 3 consists of a series of hesitations and a response cry (end of line 3, 'haH:', Goffman, 1981), signalling the effort and difficulty in repeating the target item. The turn is accompanied by a modification of the patient's gaze orientation: first he gazes up, then he lowers his gaze and finally turns his head ahead and his gaze to the front (Fig. 6 and 7). By gazing away from the therapist, he manifests a recurrent pattern in a 'searching activity' (Goodwin & Goodwin, 1986). This orientation is maintained in the following turns: with the exception of a brief glance at the therapist at line 7, the patient continues to look to the front – even when the therapist restarts the sequence by repeating the target item, thus audibly restarting the pronunciation sequence.

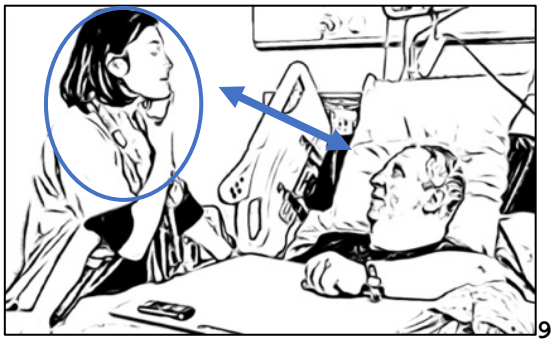
The first repetition of the target syllable (line 5) by the therapist is performed with a *high volume* of the voice (signalled in the transcript by the uppercase letters) and *emphasis* on the first phoneme (signalled by the underlining). This results in the patient's new attempt at producing the syllable. The second attempt at correcting the patient's turn (line 8) is constituted by the production of the phoneme 'B', which is first held with closed lips and without any release of air and then released in the production of the entire syllable. The therapist further highlights the pronunciation of the target phoneme by closing her lips during the pause at line 9, thus taking a visible 'labial position', while she continues to look at the patient (for a similar practice in speech-language therapy for children, see Ronkainen, 2011). The patient is nevertheless still gazing to the front. He will finally turn towards her after his new response at line 10, a response that results in the perseveration of the phoneme 'K'. Therefore, the patient does not succeed in producing the linguistic item, despite the multiple repetitions of the syllable by the therapist and the fact that the audible production of the syllable is highly emphasised through prosodic features such as *volume and emphasis.*

In the continuation of the sequence, benefiting from a mutual gaze orientation of the patient at the end of the patient's turn (line 11), the therapist makes relevant a visual perception of the target phoneme and of its pronunciation. After having negatively evaluated the turn of the patient with a click of the tongue and a shake of the head (line 12), the therapist again positions her lips on a labial position,

this time *pointing*, with her right index finger, towards her chin and touching it (line 13, Fig. 9):

Extract 1b.

```
11          +(0.3)
   pat      +turns h,gazes tw slt-->
12 SLT      £.tsk£
   slt      £shakes head£
13          Ω⊥#(0.5)Ω
   slt=>    Ωlabial positionΩ
   slt=>    ⊥pts tw chin-->
   fig      #fig9
```



9

```
14 SLT      ⊥<((hold with closed lips,without release of air)) b>⊥
      =>    ⊥hits 2 times⊥
15          ⊥ξ(0.4)ξ(0.3)
   slt      ⊥pts-->
   pat      ξopens,closes mouthξ
16 SLT      ⊥#<((hold with closed lips,without release of air))b:>⊥BA⊥
   slt=>    ⊥.....⊥moves finger in front⊥
   fig      #fig10
```

10

17 PAT ⊥#mh (0.2) ah: (0.2) ba⊥
slt ⊥bends finger--------->⊥
fig #fig11

11

18 £⊥(0.4)£⊥
slt £nods£
slt ⊥lowers hand⊥

The pointing at line 13 allows the therapist to make salient a specific area of her face, the 'mouth', and to underline the positioning of her lips by touching her chin (for a palm-up gesture that involves the chin in teaching pronunciation of L2, see Smotrova, 2017). The pointing is then exploited for realising a double hit (line 14), which functions as a *summons* and accompanies the partial production of the target phoneme (without release of air). This allows the therapist to make relevant both the visible and audible dimension of the target item ('B') and their coordination. The therapist is then guiding the attention of the patient to a specific area of her body, 'silently' (but see the audible feature of the double hit, line 14) instructing him to see and pay attention to her mouth.

Finally, the pointing is released when the therapist produces the entire syllable at line 16: the production of the linguistic item is audibly emphasised first by the stretching of the phoneme 'B' (again, produced without release of air) and then by the high volume of the entire syllable. It is also *visually represented by the movement of the index finger* (Fig. 10), which is directed from the chin onwards: this gesture accompanies the production of the entire syllable and the opening of the mouth (that is due to the release of air and the production of the second phoneme, 'A'). At line 17, the patient shows recognition of the target phoneme with a change-of-state token (Heritage, 1984). This is followed by his subsequent *production* of the target syllable.

This first extract gives us a flavour of the deeply embodied nature of speech and language therapy and of how the interconnection between different senses, such as hearing and seeing, is managed in interaction. In particular, it shows that, during an oral activity, the therapist brings the attention of the patient to her body (i.e. the mouth) in order to make *visible* the features of the pronunciation of a selected phoneme. This is done with a *self-pointing gesture* that highlights the visual articulation of the pronunciation of the item (i.e. the position of the lips); the

gesture allows a focus, first, on the mouth, and then, by accompanying the change of lip position with a movement of the index finger from the chin onwards, further emphasises the visual perception of the target sound (see line 16). Interestingly, it is precisely the deployment of both *audible and visible resources* by the therapist that allows the patient to recognise and finally reproduce the target sound.

The possibility of using lip position as a visual representation of the pronunciation requires the patient's gaze orientation. In the selected extract, the therapist indeed points to her chin and highlights the lip position only once the patient has turned towards her. Once the patient's visual attention is secured, the *self-pointing* gesture allows focusing on a specific area of the face to make relevant a *specific visible feature* during an activity that is originally configured as mainly verbal and auditory.

The two angles provided by the two cameras allow for the capture of not only how participants mutually coordinate their actions (see the therapist's pointing gesture and facial positioning following the change in the patient's gaze direction) but also the deeply intertwined production of sounds and visual cues. In particular, the mobile camera that is positioned laterally at the level of the participants' upper bodies allows detailed scrutiny of participants' gestures and gaze orientation. The view of the camera positioned behind the therapist gives clear access to the patient's face and reproduces what is accessible to the therapist – thus giving access to her visual perceptual perspective.

5.2. Representing sounds with gestures

As observed in the previous section, the therapist can accompany the vocal production of the target linguistic item with a gesture (the index finger) that reproduces the articulatory movement occasioned by the production of the two phonemes of the syllable – thus making visible a feature of the pronunciation of the target sound. In this section, we analyse an instance in which several features of pronunciation are emphasised with several representative gestures, that is, gestures that depict the target element with different techniques of representation (Kendon, 2004).

This time, the speech-language therapy takes place in a rehabilitation clinic. The participants are sitting one in front of the other and are involved in a picture-naming activity. The patient has recognised the referent of the card and is trying to produce the word *'soleil'* (sun): after multiple tries (some turns are omitted here), he manages to pronounce the word correctly (line 1) but then shows the incapacity to repeat it. The therapist continues to monitor the patient's productions, as shown by his nodding and pointing towards the patient when the patient correctly pronounces the target word. Following a new, unsuccessful repetition at the end of line 1, the therapist comes in to help and utters the target word by *segmenting* it into two syllables (lines 2, 4):

Extract 2a[3]

```
1  PAT   >le< s::o- (.) -ie- (0.2) s:oleil £⊥(0.3)£ s+:oie+(.)
         >>mid-distant gaze-->                  -->+lks slt+gazes down->
   slt   >>lks pat-->
   slt                                        £nods£
   slt                                         ⊥pts tw pat-->
2  SLT   ⊥#s::[o:
   slt   ⊥cirle wt fingers->
   fig    #fig12
```

12

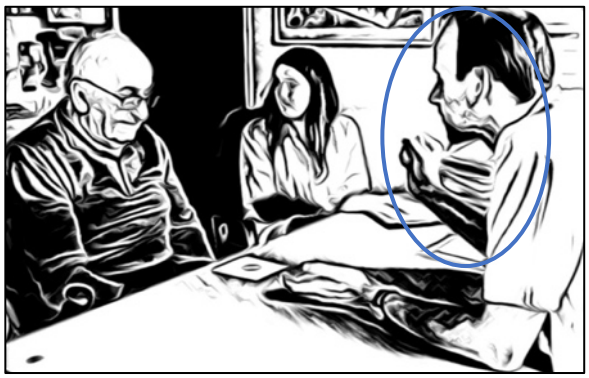
```
3  PAT          [so+ie(l)=
                -->+gazes up-->
4  SLT   =⊥#L:[:ei⊥l⊥
5  PAT          [s:o- +H:
   slt    ⊥index up,vert mouv⊥,,,⊥
   pat             -->+gazes down-->
   fig    #fig13
```

13

[3] The woman in the picture is an apprentice who is assisting in the session.

```
6          %#(0.7)
   slt     %lowers head,lks pat-->
   fig     #fig14
7  PAT     soieie (0.2)⊥#soie⊥
           -->lks down-->
   slt                  ⊥raises hand,fist gesture-->
   fig                   #fig15
8          ξ(0.2)ξ
   pat     ξbites lipsξ
```

14

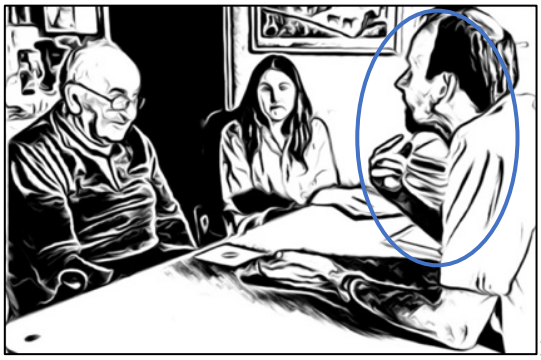


15

The turns of the therapist are audibly emphasised by a stretching of the phonemes and visually by the realisation of two gestures: the first one is a circle/ring realised with the thumb and the index finger (Fig. 12). It visually represents, and with iconicity, the closing of the mouth occasioned by the 'O' vowel. It then evolves into a vertical movement realised with the index finger (Fig. 13), which accompanies the production of the phoneme 'L', produced with emphasis with the stretching and the two vowels ('EI'). The vertical trajectory of the index finger suggests a visual representation of the rising tone, as well as a possible reproduction of the movement of the tongue (going upwards). Note that the stretching of the vocal sound is deeply synchronised with the duration of the gesture (even adjusted to it). The gesture is withdrawn during the production of the final consonant ('L'). The patient responds in overlap with an attempt at producing the entire word (lines 3, 5, 7) rather than syllable after syllable. This shows a practical problem of coordination and segmentation of the units to be repeated – which is a pattern observed also in L2 pronunciation sequences (Merlino, 2014). The patient also continues to look down during the production of the turn of the therapist at line 2, raises his gaze quickly, and then lowers it again, while trying to continue the pronunciation of the word. This 'private' activity, in which the patient shows perseveration in the production of the wrong phoneme, is interrupted gesturally by the therapist with a 'fist' stopping gesture (Fig. 15). The therapist then explicitly invites the patient to restart the sequence, redoing it collaboratively and with a slower tempo:

Extract 2b.

```
9   SLT    on va le faire >un tout p'tit peu< plus
           we are gonna do it a little bit
10         doucement >m'sieur ruban<=
           slower mister ruban
11  PAT    =+oui=
           yes
           +raises gaze-->
12  SLT    regardez
           look
13         +⊥(0.2)
    pat    +lks slt-->
    slt    ⊥.....
14  PAT    aH:ouaisH:
           oh yes
15         ⊥#(0.2)
    slt    ⊥pts chin-->
    fig    #fig16
```



16

After announcing a restart of the sequence (lines 9–10), the therapist invites the patient to look at him (‘regardez’, line 12): once he has obtained the patient’s gaze, the therapist points with his index finger towards his own chin. The self-pointing, as in extract 1, allows the focussing of the patient’s attention on the therapist’s mouth (Fig. 16). The view of one of the two cameras allows the coordination of the participants and reciprocal adjustments to be scrutinised. Positioned at the level of their upper bodies, the camera gives access particularly to the patient’s face and to the lateral side of the therapist’s body: this allows the capture of the head orientation and the right-hand gesture. Once the target of the visual attention is established, the therapist again repeats the word, segmenting it into two syllables (line 16):

Extract 2c.

```
16 SLT   ⊥#s::⊥#:SO::          ⊥#+(0.6)     ⊥#L:#:eil⊥
   slt   ⊥pts ⊥forward,ring g⊥pts tw face⊥forward index finger⊥
   pat                      -->+lowers gaze,lks slt's finger->
   fig   #fig17#fig18          #fig19       #fig20#fig21
```

```
17                 18              19              20          21
17       +(0.2)⊥
   slt   ,,,,,,⊥
   pat   +gazes down-->>
18 PAT   .h: (0.3) un(e) s::olei-i(e)H:
19       ⊥£(0.2)£
   slt   ⊥.....
   slt   £nods£
20 SLT   o:k⊥é.⊥
         okay
   slt   ...⊥places following card⊥
```

The turn is produced with a strong prosodic emphasis and is accompanied, again, by a precise gestural and visual representation of the vocal sounds: the camera positioned behind the patient, angled down above the participants' heads, captures the therapist's face (and mouth) and gives frontal access to the deployed gesture. The index finger pointing first accompanies the production of the stretched 'S' (Fig. 17). It is then transformed into a 'ring' gesture (while producing the 'O' vowel; Fig. 18) and then, in the following pause, converted again into a vertical index finger pointing gesture (Fig. 19). As observed in extract 1, the pointing is released with a gesture that goes from the chin onwards and that accompanies the production of the following syllable (LEIL), with, again, an emphasised and stretched production of the phoneme 'L' (Fig. 20–21). The trajectory designed by the gesture visually represents the duration of the syllable. It both allows the maintenance of a focus on the mouth area as well as embodies the duration of the sound and, plausibly, the upward movement of the tongue. The patient responds at line 18, reproducing the target item; the therapist accepts the answer and moves on with the activity (lines 19–20).

To summarise, the oral production of the target word is accompanied and synchronised with precise and clear gestures that not only direct the focus to the therapist's mouth but that also reproduce features of the mouth and tongue's

movements and of the target sounds. Pronunciation is then emphasised by the therapist with 1) the use of segmentation of the word into syllables and specific prosodic features, 2) the focus on the position of his lips, and 3) *gestures* that reproduce and visually embody the articulatory movements and suprasegmental features such as duration. As the extract shows, this embodied experience of pronunciation becomes a resource for the patient only when explicitly framed with verbal (summons and directives) and visual (self-pointing) resources that allow the patient's visual attention to orient towards the therapist. Indeed, the same gestures realised by the therapist in the first part of the extract (extract 2a) were not taken into consideration by the patient, who was not gazing at the therapist in that precise moment. It is then part of the therapist's work to make relevant and accountable an orientation towards the visual dimension of the oral speaking activity – particularly, to *instruct about a vision of his body as an instructional tool* in the pronunciation activity.

5.3. Making vision 'accountable' through verbal and haptic resources

In this last section, we analyse a further occurrence of the therapist's use of his body for facilitating the oral production of the patient. In this case, the therapist needs to instruct the patient about the relevance of the visual support of the lip position the patient is offering. This is occasioned by the presence of different foci of visual attention, which are due to the contextual configuration of the setting and of the task.

The participants, who are at the rehabilitation centre, are involved in a reading aloud activity. As a support they are using a computer on which the target word is displayed, which is also supported by an image of the referent (a *'mouchoir'*, that is, a 'tissue'). The patient is looking at the screen, while the therapist is looking at the patient, monitoring his verbal and visual conduct, as we can see in the following image:

Figure 22.

The image has been taken from the fixed camera positioned behind the patient, slightly above the participants' heads: this view allows access to the participants' upper bodies and faces, as well as to the material environment. It particularly emphasises the visual elements that are accessible to the patient: the therapist's face and upper body and the computer screen.

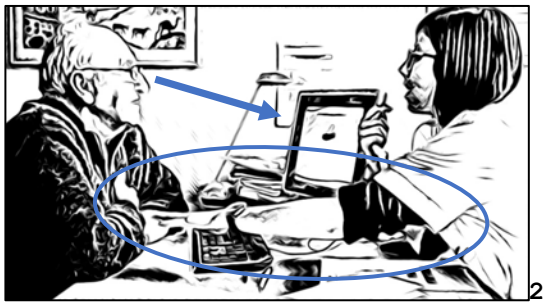
The patient has tried to pronounce the target word several times but without success (this part is omitted in the following transcript). After the production of a further attempt (line 1), he looks away from the computer by orienting his gaze to the front but producing a mid-distance gaze and not looking at the therapist (again showing a recurrent pattern in a searching activity). This orientation accompanies the hesitation in line 1 and a further interrupted attempt at pronouncing the word. During the following pause, the patient starts to rub his left eye. The therapist then recalls his attention, first by calling him (line 3) and then inviting him to 'look' (*'regardez un peu'*, line 5). However, while the therapist in the following pause (line 6) starts to point towards her chin, thus preparing to show her lip position (by clarifying retrospectively the meaning of the directive), the patient instead raises his head and looks at the screen (Fig. 23). This conduct is visible in the other view of the camera manipulated by the researcher and positioned laterally at the level of the participants' heads.

Extract 3a.

```
1  PAT   .h meuch:      +(0.2) °euh° (0.3) m::-
   pat   >>lks screen--+gazes in front/up-->
   slt   >>arm wth elbow on the table-->
2        +Δ(0.5)
   pat   +gazes/h down-->
   pat   Δrubs his eye-->
3  SLT   .tsk monsieur ruban,
         .tsk mister ruban
4        (0.4)
5  SLT   regardez un peu
         take a look
6        (0.2)Δ+⊥#(0.2)
   pat      -->+raises gaze,lks at screen-->
   pat   -----Δ
   slt   ⊥......⊥pts tws chin wth right index-->
   fig           #fig23
```

23

```
7  PAT   .h:: °°ah oui°°
                oh yes
8  SLT   re- regardez £#re+gardez+£⊥#l'ouverture£⊥de mes lèvres⊥
         1- look        look          at the opening of my lips
   slt                £approaches LH to pat's arm£,,,£
   slt                         -->⊥slight circle wt RH⊥points⊥
   pat               --->+.....+turns tw/lks slt-->1.15
   fig                 #fig24       #fig25
```



24

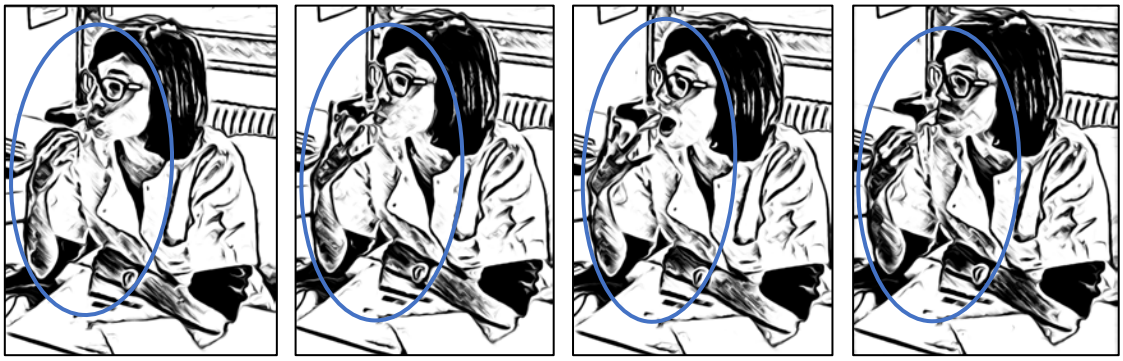
25

It is this visual conduct of the patient that occasions an explicit comment from the therapist. She repeats the directive, 'look' as a summons but does not receive a visual response from the patient. The therapist then utters the directive again while reaching out her left *hand* to the patient's *arm* (Fig. 24), thus projecting a form of touch (for uses of touch as an attention-getting device in this context, see Merlino, 2020). The trajectory of the hand is modified following the change in

gaze direction by the patient: indeed, the patient responds by turning his head towards the therapist, who, at this precise moment, withdraws her left hand and marks, with her right index finger, a slight circular movement around her lips. This circular movement accompanies the specification of the object of the recalled visual attention: the opening of her lips (*'ouverture de mes lèvres'*, line 8). Once she has secured the patient's visual attention, and indicated what exactly he is to look at, the therapist performs the pronunciation of the target word, which is emphasised both audibly and visually.

Extract 3b.

1 SLT Ω⊥.h::Ω⊥#m::[ou:⊥#CH#OI::[:⊥#R:⊥
2 PAT [mou [mou::choi-
slt Ωround lipsΩ
slt ⊥grasps hand⊥....⊥opens hand⊥,,⊥
fig #fig26 #fig27 #fig28 #fig29



26 27 28 29

3 (0.3)
4 PAT mou:ch[oir
5 SLT [⊥VOILA⊥
that's it
slt ⊥pts tw pat⊥

The preparatory positioning of the lips (round position on the target phoneme 'M') is accompanied by the grasped right hand turned in the direction of her face (Fig. 26). The first syllable of the word is pronounced with a stretching of the sounds and a movement of the hand, which is progressively opened (Fig. 27). The production of part of the second syllable ('CHOI') is characterised by a high volume of the voice and stretching. These prosodic audible features are accompanied by a further opening of the hand (Fig. 28); indeed, the production of the vowel sounds is emphasised not only by the opening of the mouth but also by its iconic representation with the hand. The emphasised opening of the mouth results also in a movement of the head and a subsequent widening of the eyes and raising of the eyebrows. The gesture is withdrawn when the therapist produces the last phoneme of the target word ('R'). Once more, the therapist's audible and visible representation of the pronunciation succeeds in helping the

patient to recognise and repeat the target sounds as the following turn shows (line 12).

A cluster of resources is deployed by the therapist in order to assist the patient in accomplishing the task: the facial expression (mouth, eyes, eyebrows), the deictic pointing gesture and the iconic representative gesture. All these embodied resources are *coordinated* with the vocal productions. Their accountability is verbally framed by the therapist so as to handle an absence of visual orientation of the patient and is further reinforced by a gesture that projects touching as a form of summons[4] (see also Merlino, 2020). The absence of visual orientation from the patient confirms the fact that looking at the therapist's face and at her mouth during a reading activity is not self-evident. The contextual configuration of the setting in this case even favours an absence of mutual gaze and an orientation towards another focus of attention, that is, the computer. The therapist must then guide and instruct the patient about the relevance of the use of her body's visual representation and the use of visual cues for the accomplishment of the task. Note that the possibility of describing this visual conduct and representation is afforded by the camera angles. While the lateral camera allows for the monitoring of the participants' mutual conduct and orientation, the camera positioned behind the patient allows the detailed capture of the therapist's facial expressions – showing clearly what is accessible by the patient: that is, his visual perceptual perspective.

6. Conclusion

In this paper, I have highlighted some specificities of visual practices used in speech-language therapy for the treatment of aphasia and, more particularly, in activities devoted to the production of linguistic items. By describing the visual resources used by the therapist in order to correct and instruct the patient about articulatory features of the speech sounds he produced, I showed how, in this context, pronunciation was experienced as both an audible and visible phenomenon. The therapists used their bodies as an instructional tool in order to make pronunciation visible, recognisable and repeatable. They proposed forms of sensorial access to language to the patients that were based on hearing and seeing and that were conveyed by verbal and gestural resources.

I highlighted the preparatory work done by therapists for bringing patients' attention to the therapists' own face for patients to see, thus making relevant, in a speech activity, the visual dimension. Through self-pointing gestures, realised with the index finger (extracts 1, 2) or with the entire hand (extract 3) in the direction of the chin, the therapists instructed the patients to focus on their mouth.

[4] Cf. Cekaite (2016) for the use of haptic resources in managing children's participation and inappropriate displays of attention.

These gestures could also be accompanied and verbally framed by directives, verbal and haptic summons and instructions that sustained the accountability of the practice. The production of the target linguistic item was realised not only with a focus on the mouth area (through the pointing gesture) but also with the deployment of audible and visible resources that emphasised features of the pronunciation: prosody (volume, duration, emphasis); representative gestures, such as thumb and index circles; index finger horizontal and vertical movements; and opening of the hand. All these gestures iconically represented (and doubled) features of the position of the mouth and tongue (such as closing or opening of the mouth or raising of the tongue). In other words, the gestures made salient the specific, articulatory features of sound production.

The use of all these resources showed that the production of linguistic items is perceived and practiced as a deeply embodied experience, and that the instructional work done by the therapists to assist the patients in exercising speech and language therapy highly relies on embodied resources (see also the 'stopping gestures' realised to stop 'perseverations' of the patient in producing the same incorrect phoneme, extract 2) and on the coordination of audible and visible cues. From this perspective, the production of linguistic items and pronunciation is treated by participants as a sensory experience and an 'intersubjectively and intercorporally organized accountable practice' (Mondada, 2019, p. 48). The accountability of the practice, nevertheless, had to be worked out by participants, as the therapists needed to drive the attention of the patients to the face, showing relevant visual details and letting them 'see' these details in order to benefit from them. The analysis of extracts 2 and 3 showed that, on the one hand, if such visible cues are deployed when the patient is not seeing them, they are not effective. On the other hand, the therapist sometimes needs to explicitly call for the patient's visual attention to the lips and to teach a practice that seems specific to instructional settings (for L2 classroom interactions, see Smotrova, 2017). The use of touch as a summons in extract 3 (see also Merlino, 2020) invites reflection on how making sensory practices accountable makes participants rely on other senses. Finally, sensory practices related to senses, such as hearing, seeing and touching, seem to play a crucial role in the process of speech and language therapy and thus deserve further investigation.

The detailed analyses of participants' visual and auditory conduct were possible thanks to detailed transcription of these resources, which, in turn, were possible thanks to the type of audio and video recordings realised during the fieldwork. In particular, I underlined the importance of using a voice recorder for a good quality of the audio and two cameras in order to document, in a complementary way, both participants' perspectives: the two distinctive views offered by the two cameras, favoured also by local adjustments realised by the researcher using the mobile camera, allowed the capture of details of participants' upper bodies and faces and to document the sensorial practices described in this paper.

Transcription conventions

For the verbal resources:

http://icar.cnrs.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf

For the multimodal resources:

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

Acknowledgements

The author wishes to thank two anonymous reviewers and the editors for their insightful comments on a previous version of the paper.

References

Beeke, Suzanne; Beckley, Firle; Johnson, Fiona; Heilemann, Claudia; Edwards, Susan; Maxim, Jane & Best, Wendy (2015). Conversation focused aphasia therapy: Investigating the adoption of strategies by people with agrammatism. *Aphasiology, 29*(3), 355–377.

Cekaite, Asta (2016). Touch as social control: Haptic organization of attention in adult–child interactions. *Journal of Pragmatics, 92*, 30-42.

Cekaite, Asta & Mondada, Lorenza (eds.) (2020). *Touch in Social Interaction: Touch, Language, and Body*. New York: Routledge.

Dunn, Isabelle (2010). *The effects of multimodality cueing on lexical retrieval in aphasic speakers.* Wayne, NJ: The William Paterson University of New Jersey.

Egbert, Maria & Deppermann, Arnulf (2012). *Hearing aids communication: Integrating social interaction, audiology and user centered design to improve communication with hearing loss and hearing technologies*. Mannheim: Verlag für Gesprächsforschung.

Ferguson, Alison & Armstrong, Elizabeth (2004). Reflections on speech-language therapists' talk: Implications for clinical practice and education. *International Journal of Language and Communication Disorders, 39*(4), 469–477; discussion 477–480.

Fridriksson, Julius; Baker, Julie M.; Whiteside, Janet; Eoute Jr., David; Moser, Dana; Vesselinov, Roumen & Rorden, Chris (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke, 40*(3), 853–858.

Goffman, Erving (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.

Goodwin, Charles (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Goodwin, Charles (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Goodwin, Charles (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522.

Goodwin, Charles (2003). Pointing as situated practice. In K. Sotaro (ed.), *Pointing: Where language, culture and cognition meet* (pp. 217–241). Mahwah, NJ: Lawrence Erlbaum.

Goodwin, Marjorie Harness (2017). Haptic sociality. In C. Meyer, J. Streeck, & J. S, Jordan (eds.), *Intercorporeality: Emerging socialities in interaction* (pp. 73–102). Cambridge, MA: Oxford University Press.

Goodwin, Marjorie Harness & Cekaite, Asta (2018). *Embodied Family Choreography. Practices of Control, Care, and Mundane Creativity*. New York: Routledge.

Goodwin, Marjorie Harness & Goodwin, Charles (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica: Journal of the International Association for Semiotic Studies/Revue de l'Association Internationale de Sémiotique, 62*(1–2), 51–76.

Gullberg, Marianne (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *International Review of Applied Linguistics in Language Teaching, 44*(2), 103–124.

Heath, Christian (1986). *Body movement and speech in medical interaction*. Cambridge: Cambridge University Press.

Heath, Christian (2018). Embodying Action: Gaze, Mutual Gaze and the Revelation of Signs and Symptoms during the Medical Consultation. In D. Favareau (ed.), *Co-operative engagements in intertwined semiosis: Essays in honour of Charles Goodwin* (pp. 164–177). Tartu: University of Tartu Press.

Heath, Christian; Hindmarsh, Jon & Luff, Paul (2010). *Video in qualitative research*. London: Sage Publications.

Heritage, John (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (eds.), *Structures of social*

action: Studies in conversation analysis (pp. 299–345). Cambridge: Cambridge University Press.

Horton, Simon & Byng, Sally (2000). Examining interaction in language therapy. *International Journal of Language and Communication Disorders, 35*(3), 355–375.

Jefferson, Gail (2017). *Repairing the Broken Surface of Talk: Managing Problems in Speaking, Hearing, and Understanding in Conversation*. Oxford: Oxford University Press.

Keevallik, Leelo & Ogden, Richard (2020). Sounds on the Margins of Language at the Heart of Interaction. *Research on Language and Social Interaction, 53*(1), 1–18.

Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge, MA: Cambridge University Press.

Kidwell, Mardi (2005). Gaze as social control: How very young children differentiate “the look” from a “mere look” by their adult caregivers. *Research on Language and Social Interaction, 38*(4), 417—-449.

Kidwell, Mardi & Zimmerman, Don H. (2007). Joint attention as action. *Journal of Pragmatics, 39*(3), 592–611.

Klippi, Anu (2015). Pointing as an embodied practice in aphasic interaction. *Aphasiology, 29*(3), 337–354.

Klippi, Anu & Ahopalo, Liisa (2008). The interplay between verbal and non-verbal behaviour in aphasic word search in conversation. In A. Klippi & K. Launonen (eds.), *Research in Logopedics: Speech and language therapy in Finland* (pp.146–171). Clevedon: Multilingual matters.

Laakso, Minna (2015). Collaborative participation in aphasic word searching: Comparison between significant others and speech and language therapists. *Aphasiology, 29*(3), 269–290.

Laakso, Minna & Klippi, Anu (2010). A closer look at the ‘hint and guess’ sequences in aphasic conversation. *Aphasiology, 13*(4–5), 345–363.

LaBonte, Andrew; Hindmarsh, Jon & vom Lehn, Dirk (2021). Data Collection at Height: Embodied Competence, Multisensoriality and Video-based Research in an Extreme Context of Work. *Social Interaction. Video-Based Studies of Human Sociality, 4*(3).

Lazaraton, Anne (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning, 54*(1), 79–117.

Liberman, Kenneth (2018). Objectivation practices. *Social Interaction. Video-based studies of human sociality, 1*(2). https://doi.org/10.7146/si.v1i2.110037

Luff, Paul & Heath, Christian (2012). Some 'technical challenges' of video analysis: Social actions, objects, material realities and the problems of perspective. *Qualitative Research, 12*(3), 255–279.

Merlino, Sara (2014). Singing in "another" language: How pronunciation matters in the organisation of choral rehearsals. *Social Semiotics, 24*(4), 420–445.

Merlino, Sara (2017). Initiatives topicales du client aphasique au cours de séances de rééducation: Pratiques interactionnelles et enjeux identitaires. In S. Keel & L. Mondada (eds.), *Participation et asymétries dans l'interaction institutionnelle* (pp. 53–94). Paris: L'Harmattan.

Merlino, Sara (2020). Professional touch in speech and language therapy for the treatment of post-stroke aphasia. In A. Cekaite & L. Mondada (eds.), *Touch in social interaction: Touch, language and body* (pp. 197–223). London and New York: Routledge.

Merlino, Sara (2021). Haptics and emotions in speech and language therapy sessions for people with post-stroke aphasia. In J. S. Robles & A. Weatherall (Eds.), *How Emotions are Made in Talk* (pp. 233-262). Amsterdam: John Benjamins.

Merlino, Sara; Mondada, Lorenza & Söderström, Ola (forth.). Walking through the city soundscape. An audio-visual analysis of sensory experience for people with psychosis. *Visual Communication.*

Mondada, Lorenza (2006). Video recording as the reflexive preservation and configuration of phenomenal features for analysis. *Video analysis*, 51–68.

Mondada, Lorenza (2014). Instructions in the operating room: How the surgeon directs their assistant's hands. *Discourse Studies, 16*(2), 131–161.

Mondada, Lorenza (2018a). Visual practices: video studies, multimodality and multisensoriality. In Favareau, D. (ed), *Co-operative Engagements in Intertwined Semiosis: Essays in Honour of Charles Goodwin* (pp. 304-325), Tartu: University of Tartu Press.

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

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

Mondada, Lorenza (forth.). *Sensing in Social Interaction. The taste for cheese in gourmet shops*. CUP.

Mondada, L., Bouaouina, S. A., Camus, L., Gauthier, G., Svensson, H., & Tekin, B. S. (2021). The local and filmed accountability of sensorial practices: The intersubjectivity of touch as an interactional acheivement. *Social Interaction. Video-Based Studies of Human Sociality, 4*(3).

Nishizaka, Aug (2011). The embodied organization of a real-time fetus: The visible and the invisible in prenatal ultrasound examinations. *Social studies of science, 41*(3), 309–336.

Nishizaka, Aug (2014). Instructed perception in prenatal ultrasound examinations. *Discourse Studies, 16*(2), 217–246.

Parry, Ruth (2005). A video analysis of how physiotherapists communicate with patients about errors of performance: Insights for practice and policy. *Physiotherapy, 91*(4), 204–214.

Pierce, John E.; O'Halloran, Robyn; Togher, Leanne & Rose, Miranda L. (2019). What Is Meant by “Multimodal Therapy” for Aphasia? *American Journal of Speech-Language Pathology, 28*(2), 706-716.

Pilnick, Alison; Hindmarsh, Jon & Gill, Virginia Teas (2009). Beyond ‘doctor and patient’: developments in the study of healthcare interactions. *Sociology of Health and Illness, 31*(6), 787-802.

Ronkainen, Riitta Johanna (2011). Enhancing listening and imitation skills in children with cochlear implants-the use of multimodal resources in speech therapy. *Journal of Interactional Research in Communication Disorders, 2*(2), 245–269.

Rose, Miranda L. (2013). Releasing the constraints on aphasia therapy: The positive impact of gesture and multimodality treatments. *American Journal of Speech-Language Pathology*, *22(2)*, 227-239.

Rose, Miranda L. & Attard, Michelle (2011). *Multi-modality aphasia therapy (M-MAT): A procedural manual*. Melbourne, Australia: La Trobe University.

Rossano, Federico (2012). *Gaze behavior in face-to-face interaction.* Nijmegen: Radboud University Nijmegen.

Simmons-Mackie, Nina & Damico, Jack S. (1999). Social role negotiation in aphasia therapy: Competence, incompetence and conflict. *Constructing (in) competence: Disabling evaluations in clinical and social interaction, 2*, 313–341.

Smotrova, Tetyana (2017). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly, 51*(1), 59–89.

Streeck, Jürgen; Goodwin, Charles & LeBaron, Curtis, eds. (2011). *Embodied interaction: Language and the Body in the Material World*. New York: Cambridge University Press.

Stukenbrock, Anja (2014). Take the words out of my mouth: Verbal instructions as embodied practices. *Journal of Pragmatics, 65*, 80–102.

Sueyoshi, Ayano & Hardison, Debra M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661–699.

Svensson, Marcus Sanchez; Luff, Paul & Heath, Christian (2009). Embedding instruction in practice: contingency and collaboration during surgical training. *Sociology of Health and Illness, 31*(6), 889–906.

Wilkinson, Ray (2004). Reflecting on talk in speech and language therapy: Some contributions using conversation analysis. *International Journal of Language and Communication Disorders, 39*(4), 497–503; discussion 503–497.

Wilkinson, Ray (2011). Changing interactional behaviour: Using conversation analysis in intervention programmes for aphasic conversation. In Antaki C. (eds), *Applied Conversation Analysis* (pp. 32–53). London: Palgrave Macmillan.