

This is a repository copy of *Holding gestures across turns: Moments to generate shared understanding*..

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/75687/>

Version: Published Version

---

**Article:**

Sikveland, Rein Ove and Ogden, Richard Albert orcid.org/0000-0002-5315-720X (2012) Holding gestures across turns: Moments to generate shared understanding. *Gesture*. pp. 166-199. ISSN 1569-9773

<https://doi.org/10.1075/gest.12.2.03sik>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# John Benjamins Publishing Company



This is a contribution from *Gesture 12:2*

© 2012. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

# Holding gestures across turns

## Moments to generate shared understanding

Rein Ove Sikveland and Richard Ogden

University of Oxford / University of York

It is widely supposed that speakers only gesture while speaking. In this paper, we consider how participants in Norwegian conversation use gestures held beyond the end of a turn-at-talk as a way to handle issues of shared understanding. Analysis combining the techniques of conversation analysis, linguistic, phonetic and visual analysis, demonstrates how participants use and orient to such held gestures as displays of occasions where participants do not (yet) have a shared understanding. The paper discusses how understanding is explicitly brought forward in a sequence of turns, and how shared understandings are reached and marked through a combination of spoken and gestural elements. The paper emphasizes the temporal progressivity of talk, the delicate timing of speech and gesture relative to one another, and the participants' collaboration in successfully achieving and maintaining intersubjectivity.

**Keywords:** intersubjectivity, turn construction, turn-taking, enchrony, understanding, interactional gesture, gesture hold, social action

### Introduction

#### *Intersubjectivity and shared understanding*

One of the foundations of successful human interaction is that participants establish and maintain a convergent understanding of each other, and of the world they inhabit, a phenomenon known as intersubjectivity (Schegloff, 1991; Heritage, 1984). Intersubjectivity emphasizes that participants in interaction share a view of the world around them; without intersubjectivity, successful conversation is not possible. There are many dimensions of interaction through which participants in interaction create and maintain intersubjectivity, from a displayed awareness of one another, the timely production of an appropriate next turn, the use of deixis indexing a shared understanding of reference (e. g.,

Heritage, 2007; Schegloff, 1996; Clark & Wilkes-Gibbs, 1986), to practices of repair (Schegloff, 1992).

### *Understanding as a process, not a state of mind*

One of the cornerstones of intersubjectivity is shared understanding: participants in a conversation understand, for example, that they are referring to the same events or phenomena. Participants can explicitly claim to understand one another (“I understand”, “OK”, “right”), but their shared understanding is also embedded in their shared use of features such as reference and deixis, and through the tying together of turns at talk into sequences (Heritage, 1984). There are also practices available to speakers to signal perturbations in their understanding, most notably through next turn repair (Schegloff, 1992). Understanding can be seen as process or as a state of mind; as conversation analysts, we examine it in this paper as a temporally-bound achievement accomplished through (and embedded in) turns at talk, as a collaborative, ‘public’ achievement (Garfinkel, 1967) rather than as a private matter in an individual’s mind (Linell, 2009). Understanding is actively managed by participants in conversation (Mondada, 2011; Hindmarsh, Reynolds, & Dunne, 2011; Goodwin, 2000). This approach is consistent with, though in a different tradition from, that of Clark and colleagues, who likewise see understanding as a joint activity, incrementally and iteratively performed, with careful co-ordination in time (e.g., Clark & Brennan, 1991).

Our analytic focus is sequences of turns as the vehicle through which a speaker displays a lack of understanding, and through which understanding is collectively negotiated and achieved. The sequence we consider in this paper has both verbal and gestural features, which are co-ordinated with one another and oriented to by participants.

### *Speech and gesture as meaningfully co-ordinated activities*

Many studies have shown that speech and gesture are co-ordinated with each other in ways that enhance meaning (e. g., McNeill, 1992, 2005; Bavelas, 1994; Kita, 2000; Goldin-Meadow, 2003; Kendon, 2004; Streeck, 2009); and that in conversation, turn construction is multi-modal, involving gestural as well as linguistic elements (Goodwin, 1979; Goodwin & Goodwin, 1992; Mondada, 2006, 2007; Schegloff, 1984; Hayashi, 2003; Walker, 2012). Crucially, although speech and gesture are in different modalities, they both contribute to the expression of meaning (Kendon, 1980, 2000, 2004; McNeill, 1985, 1992). The conjunction of interacting semiotic systems allows interactants to produce multi-layered utterances which may index many kinds of meaning simultaneously.

There is some debate over the precise temporal alignment of gesture with speech (see, e. g., McNeill, 2005, for an overview). Gestures have been found to be temporally aligned with speech to display their association with verbal content (Goldin-Meadow, 2003; McNeill 2005; Habets et al., 2011), sometimes called a 'lexical affiliate' (Schegloff, 1984). On a more detailed level, a tight temporal relationship has been demonstrated between gesture peaks, eye-blinks and prominent syllables, synchronized around rhythmic 'pikes' (Loehr, 2007). The issue remains however, in how to define the relevant parts of a gesture (e. g., where it begins) in relation to spoken elements. In this paper we use the gesture phases as defined by Kendon (2004) to study their timing with other parts of turn-construction.

### *Gesture holds in general terms*

The particular gestural phenomenon in this paper is gesture holds: stretches of talk where a gesture is held beyond the end of spoken material in a Turn Constructional Unit (Ford, Fox, & Thompson, 2002; Selting, 2000; Sacks, Schegloff, & Jefferson, 1974). We outline our data in more precise detail in the Material and Methods section. In previous research gesture hold refers to a temporary 'freeze' of a gesture's movement surrounding a gesture's main phase ('stroke'), used to secure alignment of the gesture movement with speech (McNeill, 2005; Chui, 2005). The gesture holds investigated here qualify as what is referred to as post-stroke hold (Kita, 1993), which in general terms are used to extend the expression conveyed by the stroke (Kendon, 2004).

Some previous work on gesture holds shows that a gesture hold marks the ongoing relevance of the action of the TCU. For example, Kendon (1995) studied gestures as a question-marking feature in Southern Italian. As part of his study, he reported on some instances where the gesture continued well beyond the point where the speaker's turn finishes, and argued that this use of gesture served to make clear that what has just been said forms a question addressed at the co-participant. A related finding is provided by Mondada (2007), who focussed on the use of pointing in projecting a next turn. As part of her study she investigated an instance of a question, in which the accompanying gesture persisted until all verbal content of the gesturer's turn was complete. The pointing gesture was then held until the end of the co-participant's answer, and "stopping just before her [the gesturer's] acknowledgment" (Mondada, 2007, p.216). These studies show that gestures can have particular turn-transitional functions (see also Streeck & Hartge, 1992), in this case enhancing the relevance for a co-participant to respond (cf. Stivers & Rossano, 2010; Lerner, 2003).

### *Interactional linguistics*

Linguistic constructions arise from a mapping between form and meaning. This mapping is arbitrary but conventional, i.e., socially organized, and part of what speakers know about their language. In interactional linguistics (Couper-Kuhlen & Selting, 1996), we also see linguistic forms as expressions of interactional function, or in conversation analytic terms, action. This shifts the view of language as a product to a process-oriented ('enchronic', Enfield, 2009) view, according to which linguistic structures provide interactants with resources to handle aspects of interaction as it unfolds in time. Examples of action include the management of turn-taking, sequence management (such as marking beginnings and endings of sequences of talk, or relations between turns at talk), and aspects of repair. Typically, interactional linguistics (like conversation analysis, on which it is founded) has action as its target of inquiry, asking how interactants use language to implement actions. Action-centred accounts of linguistic form such as ours show how, e.g., syntax, lexis and phonetics work together to convey socially negotiated meanings.

In this paper, we will show how speech and gesture are co-ordinated across turns at talk to produce composite utterances that can be used to achieve and display intersubjective understanding. By producing gesture holds, interactants display a problem of understanding — such as identifying a referent, interpreting an ambiguous turn at talk, or making sense of a more complex telling. The gesture is held while work is done by the interactants to resolve this problem of understanding. The point at which such gestures are retracted coincides in time with places in the spoken components of turns where the problem of understanding is resolved to the satisfaction of both parties.

In the next section, we describe the data collection and analysis. The main part of the paper provides analysis of representative examples where gestures are held across the end of a spoken turn, and retracted as a method for displaying the accomplishment of a shared understanding.

## **Material and methods**

### *Material*

The material is a set of dyadic conversations in Norwegian, collected at the Institute of Speech, Hearing and Science, the Royal Institute of Technology (KTH) in Stockholm, Sweden. The Norwegian material is a subset of a larger audiovisual collection for Swedish ('Spontal'; see Beskow et al., 2009), and consists of four dyadic dialogues, each of 30 minutes in duration. Each recording has two parts: in the first one the participants were encouraged to chat, and in the second part

the participants were instructed to investigate and discuss the identity of some “mysterious” objects they would find in a box present in the recording studio. The conversations took place in a sound-proof recording studio, which secured studio-quality audio recording, and the video recordings were done with high-definition digital video cameras. The video cameras were set up to capture both participants from two different angles, each capturing the back/profile of one participant and the front of the other (upper body down to knees). All participants were native Norwegian speakers, from south-east Norway, in and around Oslo.

### *Data collection*

The data collection and analysis was done with the use of audiovisual annotation tool ELAN (for further information about ELAN, see <http://www.lat-mpi.eu/tools/elan/>). Our data collection is taken from the first part of the recordings, where the participants chat freely. This amounts to about 80 minutes of material. We excluded talk about the unusual objects to avoid potential complexities, e. g., in analysing the handling of physical objects along with gestural analysis.

Our data consists of gesture holds in sequences of talk where the participants display and resolve a problem of shared understanding. The sequence consists of two turns by different speakers. The first turn displays some lack of understanding, which is dealt with in the second turn: that is, the producer of the second turn displays an understanding of the first turn as having displayed a problem in understanding and addresses this problem. In third position (Schegloff, 2007), the participants display their understanding of the problem having been resolved. The core pattern in the data is exemplified in Example 1, displaying the temporal co-extension of speech and gesture (a more detailed transcription, e. g., of gesture phases, will be used in later versions of this example).

(1)

GESTURE (T)	-----	
T:	'HER i `SVERige - (0.3)=	[´J ]A okay
	<i>here in Sweden</i>	<i>yeah okay</i>
L:	=´NEI: i `NOR´ge men[er jeg;]	
	<i>no in Norway I mean</i>	
TIME	-> -> -> -> -> -> -> -> -> -> -> -> -> -> -> -> ->	

Example 1 consists of three distinct spoken turns. First Tor (T) makes a reference to a location (“here in Sweden”), and in doing so displays his understanding of something in Lars’ (L) prior talk. After a delay of 0.3sec, his understanding is

corrected by Lars (“no, in Norway I mean”) and this correction is confirmed by Tor (“yeah okay”) in overlap. This example involves Tor’s gesture hold, which is released during the third position “yes okay”.

We were interested in sequences that promoted shared understanding as an explicit issue, i.e., where understanding becomes topicalised and demonstrated rather than claimed (cf. Hindmarsh et al., 2011). Repair and understanding checks inevitably form a subset of such instances. But other types of action also promote the activity of seeking a shared understanding, for example when appealing to or testing a co-participant’s understanding. Consider Example 2 below.

(2)

01 B:           og så var det en nå i `SOM´mer ikke sant:=  
                  *and then there was one this summer right*

GESTURE (B) -----

02 B:           =i: (p) (°pth) Pe:: (0.8)[                                   [(k)ja: noe           [†SÅnt?  
                  *in:                    Bej::                                   (k) yes something like that*

03 L:   [tersburg Sa[nkt?                                   [˘N:Ettopp,  
  *tersburg Sankt    exactly*

In lines 01–02 Bengt (B) tests Lars’ (L) knowledge of the Olympics, as Lars prior to the excerpt has revealed that his knowledge of this topic (and sports in general) is poor. Bengt projects a very specific response from Lars (02), namely the location of the last Olympics (*Peking*”Beijing” in 2008). Like in Example 1, Bengt’s projection in Example 2 seeks an explicit understanding from Lars, however it is not motivated by a current failure in understanding in this case, but rather it requires Lars to demonstrate that he understands by completing the word.

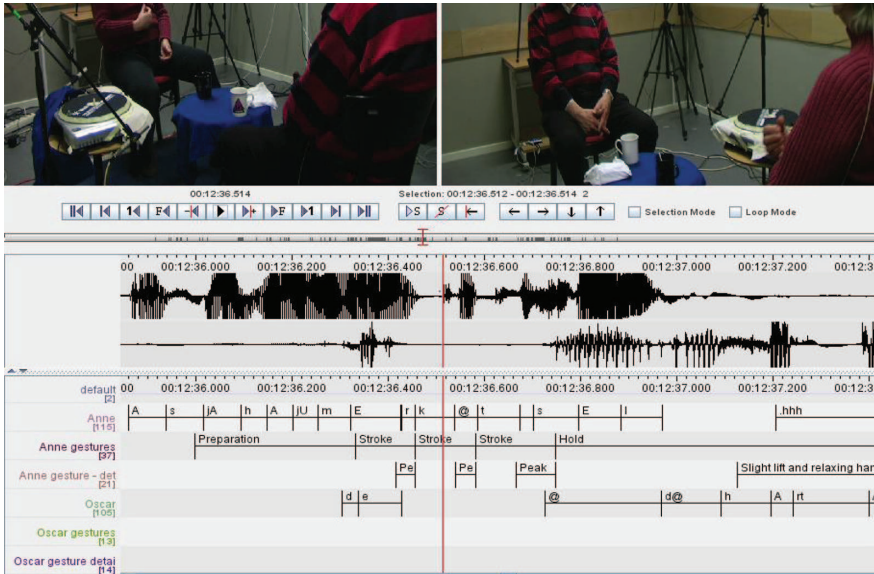
The basis for our data collection was a combination of form (gesture hold accompanying sequences similar to the ones exemplified above) and function (how the extension of gesture hold displays an orientation towards the emergence of shared understanding). However we were also interested in how commonly gesture holds were associated with the action types of interest, and for that we also made a note of instances where no gesture was used.

For simplicity, the sequence in focus will be referred to as SSU — Seeking Shared Understanding.

### *Labelling of gesture and speech*

The data collection was transcribed and labelled aligning details in speech, manual gesture (when present), and the observed presence/absence of mutual gaze. There were no restrictions regarding type of gesture, e. g., iconic, indexical, metaphoric





**Figure 1.** Illustration of ELAN annotation. Speech and gesture segments are labelled on separate tiers.

(see, e.g., McNeill, 1992), as long as they could be analysed as a relevant part of the action and meaning-making in progress.

Gesture and speech were annotated independently, and represented on separate tiers (using ELAN). Gesture annotation was done without sound, and speech annotation with no video. This was done to prevent annotation being influenced by expectations.

Gesture annotation was conducted based on a frame by frame analysis, as a means of accurately determining the timings of gesture with speech. The annotations were segmented into phases for consistency: the preparation stage, the stroke, post-stroke hold and the release of gesture (see Kendon, 1980, 2004). A gesture was labelled as held when the gesture movement came to a halt following the stroke peak. The ELAN annotation is illustrated in Figure 1.

As we were interested in the overall distribution of gesture hold across turns in our data we also collected instances of gesture hold irrespective of interactional sequence. These excluded speaker transitions formed only by short recipient responses.

### *Transcription*

For transcription of speech data, the GAT2 (Gesprächsanalytisches Transkriptionssystem 2), as developed by Selting et al. (2009), was used. GAT2

uses conventional orthographic forms, combined with a representation of some phonetic and prosodic features, including glottalisation, phrasal prominence, speech rate, loudness, and pitch movement in accented syllables and at intonational phrase boundaries (see list in appendix). The gestural transcription is based on conventions used in Kendon (2004). Other features of visible action, such as direction of gaze, are also superimposed on the verbal elements, however more simplistically than the gestures.

### Distribution of gesture holds

In our dataset we identified in total 75 instances of SSUs, 25 (or 1/3) of which were accompanied by gesture hold. In comparison 41/491 overall turn-transitions were accompanied by gesture hold into the next speaker's turn. This comparison supports the notion that gesture holds are associated with SSU type actions. The remaining 16/41 occurrences of gesture hold were found in different sequential circumstances, the most common of which we termed 'incidental incomings', as illustrated in Example 3 below.

Prior to and during this except Lars (L) talks about what he observed during some violent demonstrations in Athens, which he recently visited. Bengt's (B) incoming in line 04 (*nå*/"now"), is produced in the middle of Lars' turn (03), requesting a confirmation that these events took place on this occasion (as opposed to second-hand information). Lars orients to Bengt's "now" as request for confirmation with a minimal response *ja*/"yes" in line 05, before re-initiating his turn in progress.

(3)

```

01   L:           °ptk på ´HVEr t eneste `HJØr´ne.
                °ptk on each single corner

02               <<all >gatehjørne   så> `VAR de:t – to: (eh) poli`TIer
                street corner there were      two      police-men

                /GESTURE HOLD /

03   MG(L)       ~~~~~|~~~~~
03               me:d h° °h[h ] med   skj   ^JA med ´SKJ01d o:g
                with h° °hh      with shie- yes with shields and

04   B:           [nå,]
                now
  
```

What this and other instances of incidental incomings have in common is that a current speaker's action trajectory is put on hold by incoming talk that typically

refers to previous talk rather than the currently projected trajectory. In such cases the gesturer typically holds the gesture while responding to the incoming talk, before reinitiating the gesture along with the spoken continuation/recycling of the initial action. In Example 3 the gesture hold and re-initiation is also aligned with certain phonetic events: Lars' "yes" is prosodically marked as separate from its surrounding talk, with a high, rise-falling pitch pattern and increased loudness. By virtue of the phonetic design and alignment of this item with the gesture (hold), we can see that Lars' confirmation is a multi-modally constructed TCU embedded within a TCU. Although not in main focus in this paper, these instances of gesture hold support our research claims, in that they display that a turn's projected meaning is still in progress (i.e., 'pending' completion), even when other interactional matters are currently being dealt with.

**Main analysis: illustrative cases**

Next we present detailed illustrative examples of SSUs. When presenting the cases we will refer to a sequence which is formalised in steps (1–3), according to the interactional events leading to shared understanding (see Table 1). These steps are

**Table 1.** Formalisation of the sequence of events which lead to the achievement of shared understanding, separated in three action steps (columns) and between speakers (rows). Speech and gesture are schematised as extensions over time (Action 1 — Action 3).

Speaker	Action 1	Action 2	Action 3
A			
SPEECH	-----		-----
GESTURE	-----		
CONTENT	<i>Brings an issue regarding understanding to the surface of interaction: Uses verbal resources accompanied by gesture</i>	<i>Orients to speaker B's contribution, while holding gesture</i>	<i>Displays achievement of shared understanding. Releases gesture followed by verbal response</i>
B			
SPEECH		-----	
CONTENT	<i>Displays reciprocity</i>	<i>Produces contribution to shared understanding, e.g., a confirmation</i>	<i>Displays reciprocity</i>
TIME	-> ->		

used in the transcripts to aid the reader in accessing the material, as well as illustrating one of our main analytic points.

These steps are superimposed on Example 1b as an illustration (see top line). As will be shown later on, this example contains features which complicate the schematisation somewhat.

(1b)

ACTION	1-----2-----3-----
MG(T)	~~~~***** *****(-.-)*****
T:	´HER i `SVERige- (0.3)= [´J ]A okay here in Sweden yeah okay
L:	=´NEI: i `NOR´ge men[er jeg;] no in Norway I mean

We start our analysis by presenting a case which illustrates well the sequence of verbal and gestural behaviours that we commonly find when co-participants work to achieve a shared understanding. In this case, a referent in Lars' turn production is treated as problematic: a candidate understanding of that referent is produced, accompanied by a gesture (Action 1) which is held while the candidate understanding is confirmed (Action 2). Then the confirmation of understanding is confirmed, thus marking verbally that a shared understanding has been reached, and at this point the gesture is released (Action 3). The timing of the verbal and physical behaviours co-occurs with the accomplishment of reaching a shared understanding.

The steps of the sequence presented in Table 1 are marked with numbers and arrows in the transcript. Presence of mutual gaze is inserted above line 02. In this example mutual gaze continues throughout the excerpt. In all examples, mutual gaze is present through the steps that bring forward, contribute to, and establish shared understanding (see, e.g., Stivers & Rossano, 2010 on mobilising response). This is noteworthy, because for a co-participant to orient to a held gesture, the gesture has to be visible, and perhaps also seen by the gesturer to be visible to the recipient.

(4a) TL, 14:07, "scratche"

*Lars (L) and Tor (T) are talking about attitudes towards newly developed musical instruments, and Lars argues that people in general do not regard scratching (i.e., manipulating a turntable to make sounds) as music.*

01 L: og `FORTsatt så er °hhhh er det `VELdig vanskelig å:  
and still it's °hhhh it's very difficult to

- //Mutual gaze –
- 02        å `0vertale folk << all > om at det ja men det> `ER ††jo  
           *to convince people            that            "yes but it IS*
- mus`ikk.  
           *music"*
- 03        (1.0)
- 04 T: 1-> ˚p[h ] ʔja(ʔ)å [ˈSCRAT]ˈche,=  
   *(yes) to scratch*
- 05 L:            [˚p[h]            [forʔ ]  
   *because?*
- 06 L: 2-> =(n)˘ja,-¹  
   *yes*
- 07        3-> (.)
- 08 T: 3-> ˘mm - =  
   *mm*
- 09 L:        =¹ `fordi vbm(h˚) (-) ((swallow)) det ↓`SKJØNner ˈjeg,  
                   *because                                    ((swallow)) I can understand that*
- 10        (-)
- 11 L:        d:et er ikke musikk som så veldig mange `HØRer på men;  
                   *it's not (the kind of) music a lot of people listen to but*

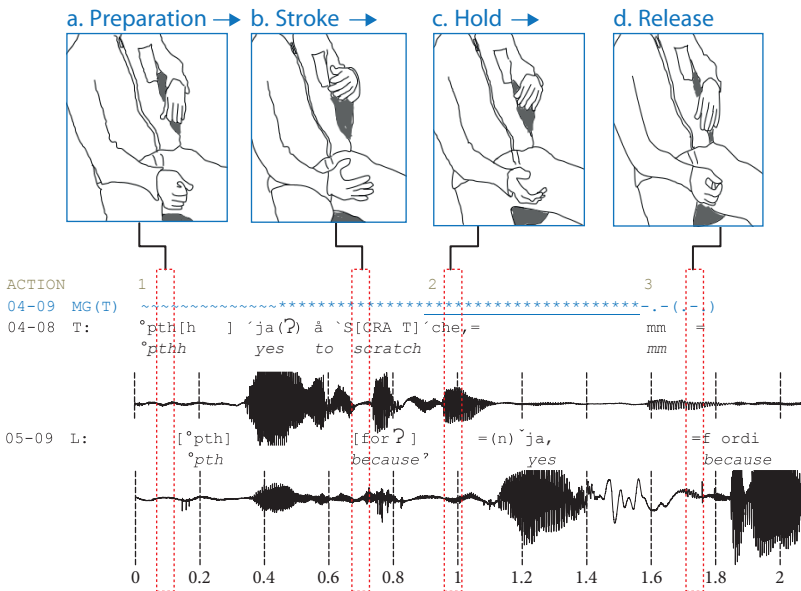
((¹ "fordi" initiating line 09 is produced as if continuing the intonation phrase in line 05))

At 01–02, Lars provides an assessment, "it is difficult..."; which as a first assessment makes relevant a second assessment from Tor, with a social preference for agreement (Pomerantz, 1984). However the 1 second gap in 03 indicates a problem on recipient Tor's part. In this inter-turn gap, Lars' gaze has moved towards Tor (who is gazing at Lars) well before the end of 02, as Lars produces "yes but it is music". The mutual gaze continues through the gap, which maintains the relevance of a response from Tor (see, e.g., Goodwin & Goodwin, 1986; Hayashi, 2003; Streeck, 2009; Stivers & Rossano, 2010, on how mutual gaze enhances the relevance for a co-participant's response). Also, the mutual gaze occurs prior to the prosodically prominent *er jo*, which further contextualises 01–02 as designed for co-participant response.

Tor's verbal response at 04 starts with [˚p[h] sounds made as the jaw is opened in readiness to speak and an in-breath is taken, followed by "yes", and a

verb phrase, “to scratch”: this expands on the pronoun “it” in Lars’ prior turn, and serves as a candidate understanding of the referent assessed at 01–02. This turn at 04 is the start of the target sequence (Action 1), as Tor’s speech is accompanied by a manual gesture held beyond the point at which the spoken part of his turn is complete. See the annotation and illustration 4b (representing lines 04–09 in transcript 4a horizontally, in time).

(4b) TL, 14:07, “scratche” - GESTURE ANNOTATION



Lars comes in in overlap with Tor, starting with a percussive and an in-breath ( $^{\circ}$ pth), and then the first part of “because”, which is cut off with a glottal stop at the end of the first syllable (see, e.g., Jaspersen, 2002; Ogden, 2001; Local & Kelly, 1986). This is also the point in time where Tor’s candidate understanding (“to scratch”) is recognisable. Following Tor’s candidate, Lars confirms it with the particle “yes”, thus confirming the referent which Tor has identified as the one Lars had intended. Tor’s gesture remains held during the whole of this period (Fig. c).

Tor releases his gesture hold (Fig. d) follows about 0.1sec after Lars’ confirmation of Tor’s candidate understanding. The release co-occurs with Tor’s particle *mm* (07), and returns to rest position on its completion. There are two parts to the withdrawal of the gesture: one with Tor’s right hand (outside brackets), the other with Tor’s left hand resting on the back of his chair (inside brackets). With *mm*, Tor signals that the referent of *det/it* has been resolved, and by withdrawing his gesture provides visible confirmation of this.

In summary, this sequence illustrates well the primary sequence identified earlier: there is a problem of understanding (Action 1); one of the co-participants produces a gesture which is held while collaborative work is done to resolve the problem (Action 2); and once a shared understanding is reached, the gesture is released (Action 3). Shared understanding in this example and many others is accomplished through (a) the sequential organisation of talk, (b) co-ordinated verbal behaviour (i.e., in spoken turns at talk), and (c) gaze which is co-ordinated with the opening and closing of the sequence in which a problem in shared understanding is identified and finally resolved.

In Example 5 we see another case involving an understanding check, which is a form of other-initiated repair (Svennevig, 2008). This time the candidate understanding is first disconfirmed and then corrected. The development of this sequence adds a few complexities compared with Example 4, but supports our general case.

(5a) TL, 11:36, "i Sverige"

*Tor and Lars are discussing attitudes towards music. In the excerpt below Lars talks in favour of the tradition of Norwegian school bands, which, unlike Swedish school-bands, have focussed more on the social aspects of music-making than musical ambition.*

- 01 L: det ˈFINS ingen: ˈh ˈST0:re krav til at det her s  
there are no ˈh great expectations for them (to)
- 02 <<a11 >`N0en gang skal> s2 n: kunne bli ˈBRA:  
to ever \* \* do great  
eller noe sånt;  
or anything
- 03 (-)
- 04 T: ˈmh=  
ˈmh
- 05 L: =man ˈGJØR ˈdet ˈbare.  
you just do it
- 06 (.)
- 07 T: 1-> <<a11 >(ja) ˈHER i> ˈSVERige –  
(yes) here in Sweden

08 2-> (-)

09L: 2/3-> <<f >'NEI: i> `NOR'ge men[er jeg;]  
no in Norway I mean

10 T: 3-> <<f >[´j JA <<a11 >o`kay.>>  
yeah okay

11 T: [mm, [((THROAT)) ]  
mm ((throat))

12 L: [°th [med ´KORPS,]  
°th with school-bands

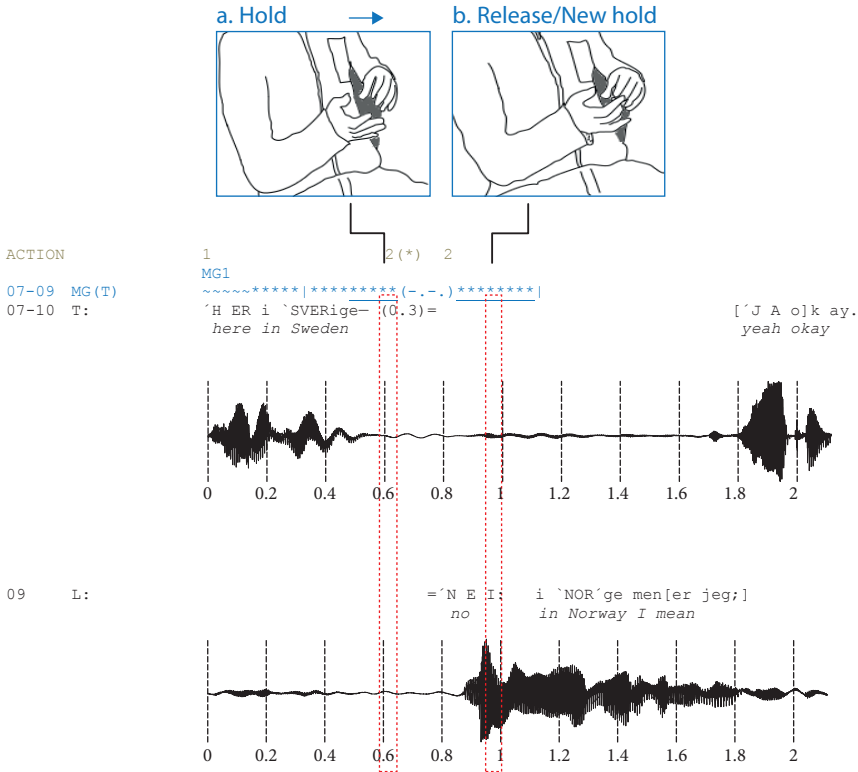
In lines 01–03, and with the increment at 05, Lars provides a turn which is prosodically, syntactically and pragmatically complete, and does not project further talk by him. Lars gazes at Tor all the way during lines 01–05. The turn therefore makes a relevant next action from Tor, such as agreement/alignment. Tor responds with an understanding check in line 07, following a brief pause (06) which displays a lack of understanding, and impeding a response such as agreement.

Alongside his turn in line 07, Tor produces a gesture which he holds into the inter-turn gap in line 08. Tor starts releasing his gesture during the gap, but then stops the release and holds it at the point when Lars initiates a response (09). The details of this development are illustrated in 5b.

Tor holds his gesture as it reaches its second stroke peak in the last syllable of *Sverige*/'Sweden' (see Fig. a). For 0.3 sec Lars does not initiate a response. 0.2 sec through this gap, Tor starts to release his manual gesture, and just after this, Lars initiates his response 'no in Norway'. Tor now stops the withdrawal of his gesture and holds it again (i.e., the handshape in Fig. b).



(5b) TL, 11:36, "i Sverige" GESTURE ANNOTATION (MG1)



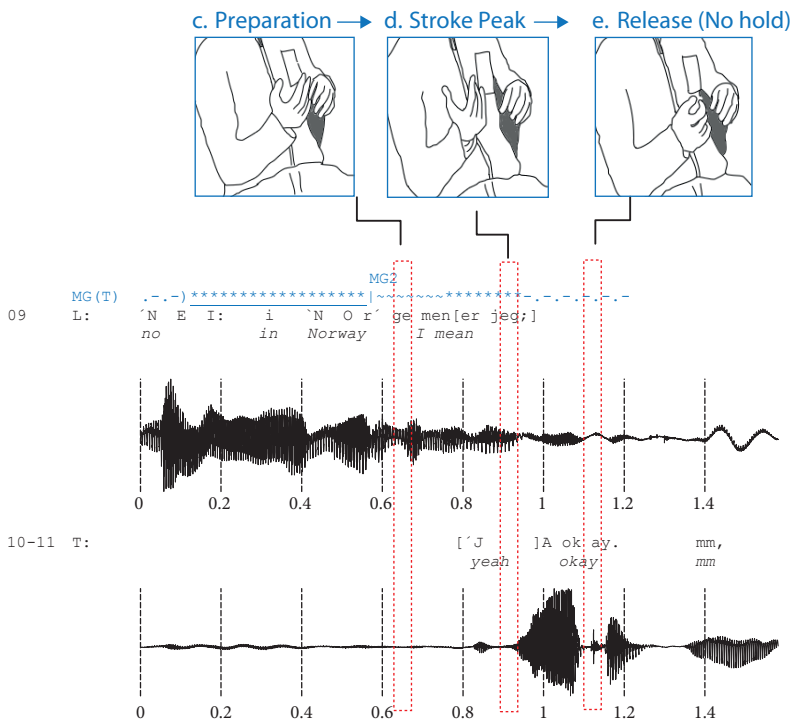
These events are of demonstrable relevance for the participants and their displayed orientation to shared understanding. First, as Tor starts releasing his gesture (Fig. b, highlighted with ‘\*’ on the Action line), Lars does not only initiate talk, he does so very abruptly, and in this way Lars displays an orientation to Tor’s release of his gesture hold. Lars’ *nei* “no” at 09 is both louder than the surrounding talk and compared to his regular volume, and this loud speech is preceded and accompanied by Lars suddenly raising his shoulders (not included in the transcript). This lends support to our claim that participants themselves treat gesture holds as relevant parts of an action in progress. As further support for this claim, Tor stops releasing his gesture and holds the ‘semi-released’ gesture in orientation to Lars’ response: the work towards achieving a shared understanding is now again in full progress.

Lars and Tor manage to achieve shared understanding within Lars’ multi-unit turn at 09. Lars’ turn consists of a rejection (“no”), and a correction (“in Norway I mean”). The correction marks “Norway” as the object in repair, by using the same

lexical/syntactic format as 07 (i.e., “in X”), whereas “I mean” marks the turn as part of a repair sequence.

As Lars produces “in Norway”, Tor produces an indexical gesture (a thumb-point) that is directed to Tor’s right and thus away from the direction of his previous flat hand gesture (Fig. a–b in 5b). This change in gesture demonstrates an orientation to Lars’ correction of Tor’s candidate understanding, that is, it co-expresses Lars’ correction of “Norway” to “Sweden” (see transcript 5c).

(5c) TL, 11:36, “i Sverige” GESTURE ANNOTATION (MG2)



Tor’s gesture hold is released and followed by a second gesture (MG2) at the offset of the of the first syllable in *noRge*. MG2 is formed as an indexical gesture using the protruding right thumb (see Fig. c–e). As Lars has already rejected Tor’s understanding, Tor may anticipate that the projected referent is another place/country, which explains Tor’s use of an indexical point away from his first gesture. Thus MG2 also contextualises the previous flat hand gesture (MG1) as indexical, and the use of MG2 facilitates the progression of shared understanding: by producing his second gesture in response to (and indeed partly in overlap with) “Norway”, Tor displays continued orientation to Lars’ correction of his *candidate* understanding, and thereby makes a visible contribution to *shared* understanding.

In this example, the point at which a shared understanding is achieved is at the point where the word *Norge* is recognisable as such. As in Example 4, this point (i.e., Action 3), is followed by a verbal confirmation (*ja okay mm*), which is partly produced in overlap with Tor's MG2. This is consistent with the relevance of timing observed in Example 4. In Example 5 the relevant gesture's (MG1) release takes the form of a new gesture (MG2), and this example further demonstrates how gestures are co-ordinated in time with the achievement of shared understanding, also when the confirmation of a candidate understanding is delayed, and needs further repair.

The gesture in Example 6 is held while a shared understanding is brought to the surface of the interaction, in this case while further on-topic talk is projected: the gesture is used here as part of a topic proffer (Schegloff, 2007), which *pre-empts* a potential problem in understanding, making it unlike our previous examples. More clearly than in previous examples, Tor's gesture in Example 6 forms a resource for making explicit his request for understanding.

(6a) TL, 7:13, "Torbjørn Thorsen"

- 01 T: 1-> torbjørn ˈTHOR`sen.  
Torbjørn Thorsen
- 02 1-> (--)
- 03 L: 2/3-> torbjørn ˈTHOR`sen ja; ˈHA[N kjenner jeg `go]dt.  
Torbjørn Thorsen (yes) I know him well
- 04 T: 3-> [mm, ]  
mm
- 05 (-)/((T:nod))
- 06 T: [mm/((nod))]
- 07 L: [((nod)) ]
- 08 T: ˚pθh hɑn hɑr jø h[ɑn gɪkk i mɪn <<f >`KLAS`Se?> ]  
˚pθh he hɑs he wɑs ɪn mɪ klɑs
- 09 L: [e]llər h˚ <<f > `KJEnte> ]:  
or h˚ knew

At 01, Tor produces only a person reference, and there are no obvious syntactic or other indications of whether or not he seeks a response from Lars, and what kind of response that might be; and indeed, Lars displays a problem in interpreting who is referred to in turn 01 by not producing an immediate response (02). In orientation to the missing response Tor produces a pointing gesture in 02 followed by a

gesture hold, marking his turn as one that seeks recognition from Lars, and which Lars then picks up on in 03.

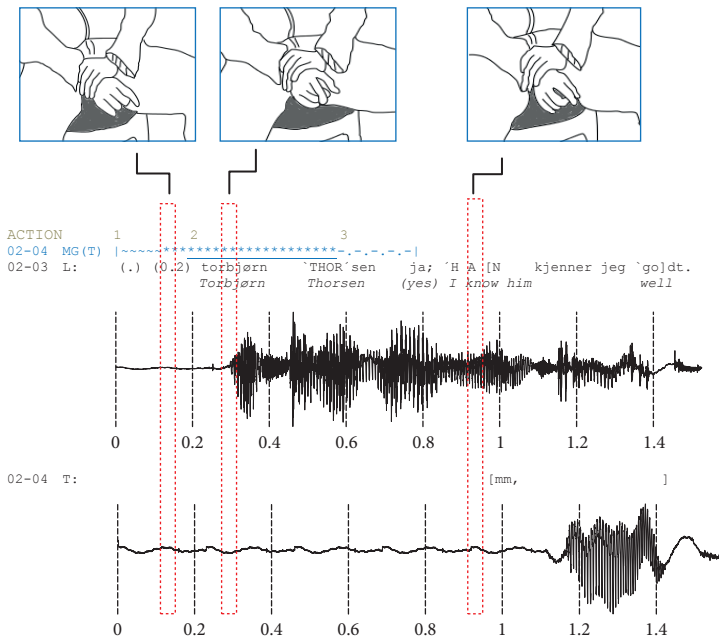
The shared action is establishing a person reference. In some circumstances participants may assume shared knowledge of a person referred to, and simply proceed to talk about him/her without any explicit demonstration of recognition from the co-participant. Such shared knowledge is not assumed here: Tor's telling cannot progress until shared recognition is established, and as such both participants display a preference for recognition over progressivity (Heritage, 2007). Several features illustrates this preference, and we will turn to these in more detail below.

Tor holds his gesture until Lars displays recognition at 03 with *I know him well*. The gestural events are illustrated in transcript 6b. Note that there is mutual gaze between the participants all the way from 01–04, enhancing the relevance for a response from Lars.

(6b) TL, 7:13, "Torbjørn Thorsen" GESTURE ANNOTATION

01 T: Torbjørn 'THOR'sen.

a. Prep/Stroke → b. Peak/Gesture hold → c. Release



Lars initiates his display of recognition soon after the initiation of Tor's gesture; this way Lars picks up on Tor's gesture as a cue to display recognition. Lars' response starts with a repeat of the name, followed by the particle *ja* "yes". This ordering

of the two TCUs serves two functions. By producing the name first, Lars orients to a display of recognition as being the relevant next action, rather than simply a go-ahead to continue (progressivity), which is what an initial *ja* might have been heard as doing. Second, compared to other instances where a recipient may repeat a word from a previous turn (e. g., when initiating repair), the following *ja* shows that the repeat was indeed a confirmation: this does not happen in repair initiations. This orientation towards a confirmed recognition is also reflected in Tor's gesture: Tor holds his gesture only for as long as it takes Lars to produce the [name + confirmation particle], at which point he releases his gesture and produces a verbal response. Thus Tor 'awaits' not only the repeat of the name but also the confirmation particle *ja* before releasing his gesture. As in the previous examples then, a gesture, its hold and timing, is of demonstrable relevance for bringing forward and resolving shared understanding.

Lars' "I know him well" (03) further qualifies Lars' earlier display of recognition: the Object — Verb — Subject word order topicalises the object "him", referring to the referent being recognised. An alternative construction for claiming recognition would have been "I know Torbjørn Thorsen well". However, the fact that Lars separates his displayed recognition into three parts: name(recognition) + confirmation particle + further qualified recognition, confirms that there is a preference for an explicit form of confirmation as well as for displayed recognition here, and arguably, this ordering benefits both of these preferences.

Our previous examples show that interactants co-ordinate the temporal alignment of their verbal and visible behaviour in order to display shared understanding. In Examples 4–6 this collaborative achievement was initiated by an individual, (i) with a candidate understanding following a displayed problem in understanding (Examples 4 and 5), and (ii) requesting a co-participant's understanding in preparation for further on-topic talk (Example 6). The resolution of understanding was achieved with the co-participant's contributions. We now turn to an example where the interactants collaboratively articulate and verbalise their shared understanding. This example provides rich evidence for the role of gestures in collaborating on intersubjective understanding.

The collaborative work gets initiated with an explicit request for shared understanding, *forstår du* "do you understand" (line 06), and is finally resolved in lines 12–13. Before 06 this issue has been presented implicitly, with the use of examples (e. g., moving from one place to another, forgetting one language in favour of another, etc.).

(7a) A0, 06:56, "språk"

*Anne (A) talks about how she had to learn several different languages as a child (her family moved from country to country), and that it was difficult for her to retain all these languages.*

- 01 A: =<sup>o</sup>h <<all >og så da jeg flyttet til> `KANa'da? (.)  
<sup>o</sup>h and then when I moved to Canada (.)
- 02 og skulle:ʔ (-) (d)ta opp `ENGelsk i'GJEN?  
and were to (-) take up on English again
- 03 <sup>o</sup>h <sup>o</sup>mh (pt) det gikk `FRYKTelig 'FORT?  
<sup>o</sup>h <sup>o</sup>mh (pt) it went terribly quick
- 04 men da <<all >var jeg nødt til å> `GL(h)EMm(h)e 'fransk,  
but then I had to forget French
- 05 0: [°pt <<breathy > [ja - > ]  
<sup>o</sup>pt yes
- 06 A: 1-> [(a) [for`ST]ÅR 'du?  
\*(you know)) do you understand
- 07 (.)
- 08 0: 2-> ja.=  
yes
- 09 A: 2-> =<sup>o</sup>h altså man k `KAN ikke: <sup>o</sup>h=  
<sup>o</sup>h you know one c- can't <sup>o</sup>h
- 10 0: 2-> =maʔ (.)[ʔ(eh) [man kan ikke] ha=  
on- (.) (uh) one can't have
- 11 A: 2-> [det [fʔ fiʔ ʔeh ]  
there a- a- uh
- 12 0: 2/3-> =´ALT: eh: ´LENGst `OPpi: [eh: ih° ] i: `HJERN´en,=  
everything uh furthest up in uh in- in the brain
- 13 A: 3-> [nʔ nei:.,]  
n- no
- 14 0: =[ehh°]  
uh
- 15 A: =[jeg ]´TROR at det finnes mennesker som `KAN ´det menʔ -  
I think there are people who can do it but-
- 16 A: ´men jeg `KAN det ik´ke,  
but I can't do it

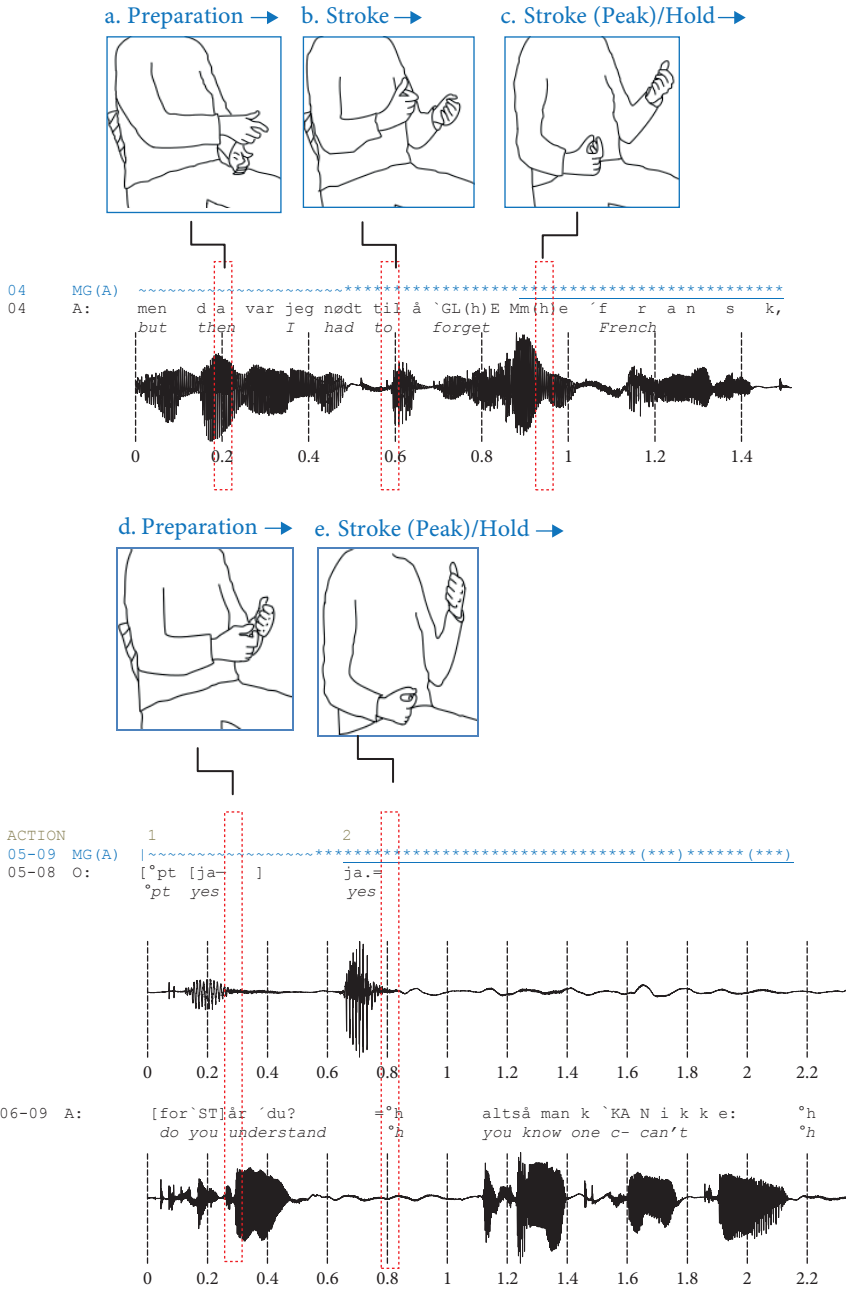
At 06 Anne initiates a sequence which explicitly changes her current project into a shared project between Anne and Oscar, as Anne produces a question-type first pair part (cf. Schegloff, 2007). The first pair part directly makes relevant a contribution (a second pair part) from Oscar, (i) by being shaped as an interrogative, (ii) by lexically addressing understanding ("understand"), and (iii) by addressing *Oscar's* understanding using the pronoun "you" (cf. Lerner, 2003). As a further enhancement of a framework for co-participation, Anne and Oscar sustain mutual gaze during the entire sequence from line 01 and until 12.

Anne's First Pair Part in 06, "do you understand", topicalises understanding, and makes relevant a confirmation of Oscar's understanding of her story so far as a story relevant to tell, and of what its implications are. In other words, it is a question of what the story is about (cf. Mandelbaum, 1989).

Anne produces a manual gesture during line 04 which she repeats when she produces "do you understand" (06). This is Anne's visual representation of her current project, which Anne and Oscar then proceed to formulate and co-construct verbally.

Anne's use of co-speech gesture is integral to the co-construction of shared understanding that she and Oscar now work towards. We will demonstrate these claims with reference to transcripts 7b-c.

(7b) A0, 06:56, "språk" GESTURE ANNOTATION 1



Anne's gesture (see Fig. a–e) is produced by both hands meeting in the space in front of her and moving in different directions. In Figure a (preparation phase)



the right hand is placed close to but above her left hand. In Figure b (stroke) Anne pulls her right hand away from the left hand while she raises the left hand upwards. This movement is continued as illustrated in Figure c. A similar gesture is then produced accompanying “do you understand” (Figures d–e): the peak of this second gesture comes following the offset of *du* “you”.

We suggest that Anne’s gesture metaphorically represents her abilities in two languages as ‘up/strong, down/weak’ (cf. Lakoff & Johnson, 2003). Her hands move apart simultaneously: while one language goes ‘up’ (i.e., is stronger), the other one goes ‘down’ (i.e., is weaker). The distance between the hands appears to be even greater the second time Anne produces this gesture, which might be a way to intensify the conflict she is trying to illustrate. The previous articulation of the gesture in conjunction with verbal material has made available a local understanding of the gesture. By repeating this gesture with the TCU “do you understand” Anne indicates visually what the shared understanding she seeks is about. Following 06 Anne holds her gesture (Fig. e).

Following “do you understand”, Oscar produces “yes” as a second pair part in 08. Anne continues to hold her gesture, a display that a simple “yes” is fitted, but not by itself sufficient (see Raymond, 2003, on the design and format of yes/no questions and answers). While continuing to hold her gesture, Anne elaborates in 09, with “you know one can’t” (“can’t” is accompanied by a small tightening of Anne’s gesture, as if further locking her hands in their positions). Oscar produces a second TCU at 10 in response to “do you understand”. This TCU recycles Anne’s talk at 09 (“one can’t”), thus treating her turn as a formulation that is also valid for him, which also marks the action of his turn as aligned with Anne’s. Anne, while still holding her gesture, abandons her verbal production in 11 as Oscar proceeds in 10/12.

At line 12, when Oscar gets to the object “everything”, he initiates a gesture which accompanies his production of “furthest up in”. This forms a significant moment in Anne’s and Oscar’s co-construction of shared meaning, illustrated in transcript 7c (Fig. f–h capture both Anne’s and Oscar’s gestures). Oscar’s gesture is formed by lifting his hands in front of his chest, palms down and fingers facing each other. Both Oscar’s hands reach the same height level accompanied by “you can’t have everything furthest up”, in this Oscar represents gesturally two things that are at the same level. In combination Oscar’s talk and gesture imply that: “you can’t have everything up there *at the same level*”.

Very soon after Oscar initiates his gesture, Anne lifts her right hand towards the same level as her left (represented by gesture annotation in brackets; see Fig. f–h), mirroring Oscar’s gesture. Anne’s gesture displays her alignment with Oscar’s turn at 12, in the way their gestures match each other in order to display shared understanding. Whereas Anne’s original gesture (in 06) represents one

language ‘up’, the other ‘down’, Oscar’s gesture in 12 represents ‘both up’. By mirroring Oscar’s gesture in 12, Anne visibly aligns herself with Oscar’s point of view.

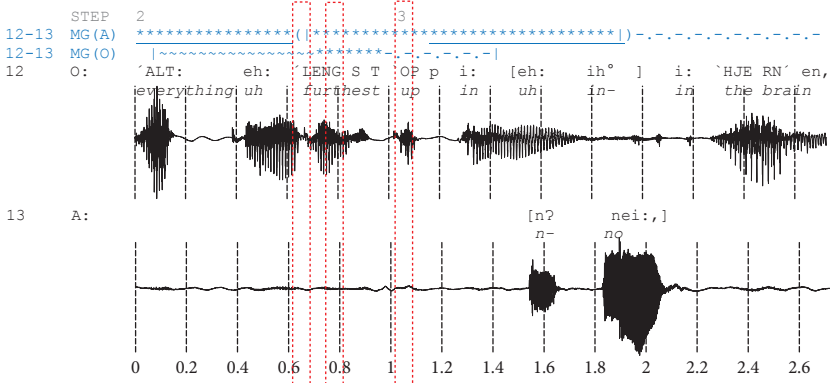
(7c) A0, 06:56, "språk" GESTURE ANNOTATION 2

ACTION STEP 2  
 10-11 MG(A) \*\*\*\*\*  
 10 O: ma? (.) [ʔ(əh) [man kan ikke] ha  
 on- (uh) one can't have  
 11 A: [det [fʔ fiʔʔəh ]  
 there a- a- uh

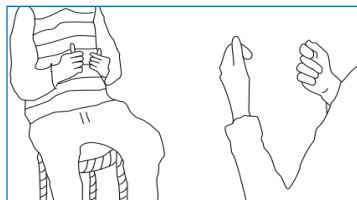
f. A: Hold  
 O: Preparation →



g. A: Stroke (Mirroring O)  
 O: Stroke →



h. A: Stroke peak  
 O: Release



Once shared understanding is available, a verbal confirmation, “no”, is given (as in British English, agreements with negatives in Norwegian are done with negatives: see Jefferson, 2002 on “no” to convey agreement in British English). There is an interesting difference between Example 7 and the earlier examples, in that Anne releases her gesture only after she has produced a verbal confirmation, whereas in the previous examples the gesture was released at, or just before, confirmation of shared understanding. We suggest that, by holding ‘Oscar’s gesture’ while confirming verbally, Anne displays that Oscar’s formulation fits with the formulation she herself projected. Thus, Anne’s extended gesture is a visual display of shared understanding at that point. Both Anne and Oscar use the same verbal and gestural metaphor, and by mirroring one another’s use of this metaphor, they have demonstrated to one another that they have understood the other to share this metaphor as well. Thus their understanding is not just subjective, but intersubjective; and visibly so (gesture mimicry as a display of shared understanding has been studied in a variety of settings, see, e.g., Kimbara, 2006; Graziano, Kendon, & Cristilli, 2011; Holler & Wilkin, 2011).

This case reinforces our contention that gestures can be held across turns at talk as a resource for speakers to display that there remains an outstanding issue of shared understanding. This case is unlike the others looked at so far, however, in that the gestures used here form a visual representation of a linguistic and cognitive metaphor, and more significantly, the metaphor is used in both linguistic and gestural terms by the co-participants to display and co-ordinate their understanding of one another’s talk.

The final example provides further evidence and support for our general claims about the uses of gesture holds. In this case, a gesture stroke (and potential hold) is relevantly avoided, as the interactants orient to mutual understanding as already being available. This example forms a deviant case (cf. Schegloff, 1968) in relation to the core sequence explored above.

Unlike the previous examples, there is no moment in Example 8 where lack of understanding comes to the surface. Quite the opposite: it becomes clear that both participants are on the same trajectory and already share an understanding. The relevant turns are highlighted at 02 and 03 in the transcript 8a.

(8a) A0, 07:50, “befinne seg”

*Oscar has been explaining how he finds it difficult to learn and use French. In 01 Anne formulates a general solution to such a challenge*

01 A:                ˘ja: h° (.) altså den ˘ENeste måten å lære seg  
                              yes h° (.)                the only way to learn

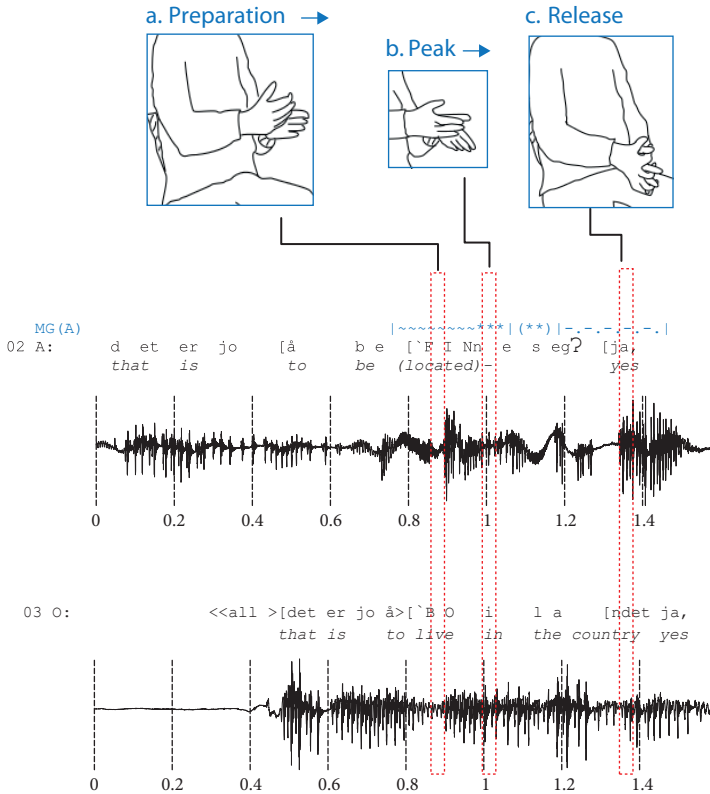
- et språk      ˘SKIKkelig –  
*a language      properly*
- 02 A: ->      det er jo [å            be [˘FINne seg? [ja,  
*that is            to            be (located)-      yes*
- 03 O: ->                    <<all >>[det er jo å>[˘BO i la      [ndet ja,  
*that is    to live in the country yes*
- 04 A:                    [˘DET er det; [det        ]=  
*that's it            it*
- 05 O:                    [javis                    [st. mm.]=  
*right                                    mm*
- 06 A:                    =<sup>o</sup>h det er jo    egentlig ^TULL    †jeg syns at det er  
*<sup>o</sup>h it's really (quite) nonsense- I think    it is*
- 07 A:                    ˘TULL            dette de ((...))  
*nonsense what they*

There are several important features that help contextualise Oscar's talk at 03 as sharing the same trajectory as Anne's. First, Oscar's 03 follows a complex NP, in which the resumptive pronoun *det*/"that" in 02 projects a comment (02) to an already introduced topic (01). Thus Oscar is likely to anticipate the projected completion of Anne's turn. Second, Oscar initiates his turn in 03 by recycling the lexis/syntax in Anne's construction in 02, *det er jo å*/"that is to". By recycling Anne's turn in this way, Oscar indicates that he is currently collaborating with her turn construction.

A third aspect of Oscar's collaboration is illustrated by how he arrives at the main verb simultaneously with Anne. Oscar achieves this by producing this turn-initiation slightly faster than Anne. The prominent syllable of Oscar's turn, *bo*/"live" (03), is time-aligned with the prominent syllable in Anne's *beFINne*/"be (located)" (02). More precisely, the release of the bilabial closure in Oscar's *bo* is simultaneous with the release of the labiodental stricture in Anne's *be**finne*. These two synchronously produced syllables are also the locations of pitch accents in their respective utterances. By temporally entraining his talk so precisely with Anne's, Oscar makes his actions recognisable as being co-constructive with Anne's actions (see also Local, 2005, on how speakers entrain to one another's talk in collaborative completions).

The peak of Anne's first gesture peak occurs 0.1sec after the onset of the prominent syllable in *beFINne*/"be (located)" (see transcript 8b).

(8b) A0, 07:50, "befinne seg" GESTURE ANNOTATION



Anne’s gesture indexically locates a place somewhere else, by thrusting her hands in a synchronised movement away from both herself and Oscar. Figures a–b present the main movement of this gesture: Fig. b shows the peak of the gesture, which is aligned with the offset of the prominent syllable in *befinne* (Anne) and *bo* (Oscar). Then, as it appears that Anne is heading for another peak she withdraws her gesture by folding her hands and bringing them towards rest position (Fig. c). This happens at the same time as she halts the production of the reflexive pronoun *seg* (as indicated in the transcript Anne’s gestural withdrawal starts in the middle of [s] in this pronoun). Her production of the vowel in this pronoun is strongly laryngealised, i.e., it shows signs of cut-off (Jasperson 2002). Also, the vowel quality does not meet its expected articulatory target (i.e., it is realised as [sæ̥] and not for example [sæɪ]).

Thus Anne halts the production of her turn both vocally and gesturally, showing an orientation to the norm of co-production of speech and gesture.

Anne’s gestural and verbal activity both stop at the point where Oscar’s turn recognisably makes the same claim that she herself is in the process of making.

Her withdrawal of her gesture and her “yes” at the end of 02 display verbally and gesturally that shared understanding has been achieved. Anne treats Oscar’s turn as a satisfactory completion of what she had started, thereby validating Oscar’s contribution. The precise timing of Anne’s withdrawal of talk and gesture confirms that a fundamental action of gesture hold is that of achieving shared understanding.

This example illustrates further how conversationalists orient to the implications of their own actions, in real time and while speaking. Example 8 supports the claim that gesture holds (and other physical display that maintain an action trajectory) exist only for as long as progress towards shared understanding is relevantly pursued. And it further illustrates conversationalists’ attention to each other’s actions while working towards shared understanding.

Overall then, our claim based on many examples and illustrated by some canonical examples here is that gesture holds visually display that progress is still being made towards a shared understanding; and withdrawal of a gesture visually displays when such a shared understanding is reached. We have focussed on sequences of turns where one speaker topicalises, or explicates, an issue in understanding addressed at a co-participant, and demonstrated how participants orient to the precise alignment of gestures with verbal material in order to secure and maintain intersubjectivity.

### **Discussion and conclusion: summary of findings**

It is known that gestures can be held across the end of a Turn Constructional Unit (Ford et al., 2002; Selting, 2000; Sacks et al., 1974). In this paper we have explored one use of gesture holds in conversation and shown that they can be used to bring a problem of shared understanding to the surface. A gesture held beyond the end of a turn is treated by a recipient as a visible display that shared understanding has not yet been achieved; it is retracted just at the point where the participants display verbally that they have reached a shared understanding. Thus the turns we have examined form ‘composite utterances’ (Enfield, 2009) as part of a longer sequence.

The linguistic, phonetic and gestural construction of an on-going turn at talk is responsive to, and shaped by, the displayed verbal and visible actions of a co-participant: that is to say, the precise linguistic or gestural shape of the turns we have examined is sensitive to the contingencies of the situation. While the linguistic and gestural forms we examine are locally constructed for the here-and-now, we have shown that there are recurrent practices in how gesture holds are used, so that they are understood and aligned with in a way that is orderly. This makes them available as resources for creating an intersubjective understanding.

Our finding is consistent with McNeill's claim (2005, p. 33), that the retraction of a gesture constitutes a display by the speaker that whatever was being dealt with gesturally is now complete:

The retraction phase, especially its end, is not without significance... It is of interest because it shows the moment at which the meaning of the gesture has been fully discharged. The hand ceases motion as it vacates its motivating force, the meaning it carried as a symbol. The end of retraction can thus show the full temporal reach of the co-expressive speech with the gesture.

While McNeill's description focuses on the gesturer, we have shown how both gesturer and recipient, across several turns at talk, orient to the gesture hold and generate the shared understanding that leads to the release of the hold.

It is fundamental to conversation analysis that analytic categories must be demonstrably real for participants themselves. By combining a multi-modal description of the data with the analytic techniques of CA, we have been able to show that gesture holds are oriented to by participants themselves as relevant for understanding one another's talk. The precise temporal alignment of the retraction of the gesture hold with verbal displays of understanding in particular indicates that models of turn design (and presumably also models of production and perception) must be able to integrate visual and verbal activities.

Language is the primary resource for conversationalists to express the meanings, and they use gesture holds as a way of marking something about their stance towards what they are saying. In other words, gesture holds provide a visible means for marking something out as 'not yet quite dealt with', and their retraction as a way of displaying (literally) that the issue has been resolved. Just as the phonetic interpretation of a sentence can alter meaning, so apparently also can co-speech gesture. Our findings therefore provide further evidence for turn-construction as an on-going, multi-modal activity (e. g., Mondada, 2007, 2011; Hayashi, 2003; Goodwin, 2000).

As we noted at the start of the paper, intersubjectivity is fundamental to human communication. The phenomenon examined in this paper provides one method for conversationalists to display visibly that a shared understanding is missing. When a co-participant attends to a gesture hold, the two conversationalists can work to achieve a shared understanding; and when it has been achieved, the gesture hold is released. Shared understanding is achieved through a process that is negotiated using speech and gesture in a co-ordinated way.

## Acknowledgements

We wish to thank the European Commission (EC) and their Marie Curie Research Training Network “Sound to Sense” (FP6-Mobility, project ref.: 35561), through which this research was funded. Within the “Sound to Sense” community we owe our gratitude to Sarah Hawkins, John Local and Sara Howard, for support and helpful input throughout this process. We wish to thank colleagues, in particular Jens Edlund, at the Institute of Speech, Hearing and Music at KTH, Stockholm, for help with collection of recording material. We further owe our gratitude to Marianne Gullberg and Adam Schembri who offered their time and effort to discuss this work with us, and gave us helpful input.

## References

- Bavelas, Janet B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction*, 27 (3), 201–221.
- Beskow, Jonas, Jens Edlund, Kjell Elenius, Kahl Hellmer, David House, & Sofie Strömbergsson (2009). Project presentation: Spontal — multimodal database of spontaneous dialog. *Proceedings of Fonetik 2008, the 11th Swedish Phonetics Conference. University of Gothenburg, Sweden, 11–13 June 2008* (pp. 190–193). Gothenburg: Reprocentralen.
- Chui, Kawai (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, 37 (6), 871–887.
- Clark, Herbert & Deanna Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition*, 22 (1), 1–39.
- Clark, Herbert & Susan E. Brennan (1991). Grounding in communication. In Lauren B. Resnick, John M. Levine, & Stephanie D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: American Psychological Association.
- Couper-Kuhlen, Elizabeth & Margret Selting (Eds.) (1996). *Prosody in conversation*. Cambridge: Cambridge University Press.
- Enfield, Nick J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- Ford, Cecilia E., Barbara A. Fox, & Sandra A. Thompson (2002). *The language of turn and sequence*. Cambridge: Cambridge University Press.
- Garfinkel, Harold (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Goldin-Meadow, Susan (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Goodwin, Charles (1979). The interactive construction of a sentence in natural conversation. In George Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Irvington.
- Goodwin, Charles (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489–1522.
- Goodwin, Charles & Marjorie H. Goodwin (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62 (1/2), 51–75.
- Goodwin, Charles & Marjorie H. Goodwin (1992). Assessments and the construction of context. In Alessandro Duranti & Charles Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 147–190). Cambridge: Cambridge University Press.



- Graziano, Maria, Adam Kendon, & Carla Cristilli (2011). Parallel gesturing in adult-child conversations. In Gail Stam & Mila Ishino (Eds.), *Integrating gestures* (pp. 89–102). Amsterdam: John Benjamins.
- Habets, Boukje, Sotaro Kita, Zeshu Shao, Asli Özyürek, & Peter Hagoort (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23 (8), 1845–1854.
- Hayashi, Makoto (2003). *Joint utterance construction in Japanese conversation*. Amsterdam: John Benjamins.
- Heritage, John (1984). A change of state token and aspects of its sequential placement. In J. Maxwell Atkinson & John Heritage (Eds.), *Structures of social action* (pp. 299–345). Cambridge: Cambridge University Press.
- Heritage, John (2007). Intersubjectivity and progressivity in person (and place) reference. In Nick J. Enfield & Tanya Stivers (Eds.), *Person reference in interaction: Linguistic, cultural and social perspectives* (pp. 255–280). Cambridge: Cambridge University Press.
- Hindmarsh, Jonathan, Patricia Reynolds, & Stephen Dunne (2011). Exhibiting understanding: The body in apprenticeship. *Journal of Pragmatics*, 43 (2), 489–503.
- Holler, Judith & Katie Wilkin (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, 35, 133–153.
- Jasperson, Robert (2002). Some linguistic aspects of closure cut-off. In Cecilia E. Ford, Barbara A. Fox, & Sandra A. Thompson (Eds.), *The language of turn and sequence* (pp. 257–286). New York: Oxford University Press.
- Jefferson, Gail (2002). Is “no” an acknowledgment token? Comparing American and British uses of (+)/(-) tokens. *Journal of pragmatics*, 34 (10/11), 1345–1383.
- Kendon, Adam (1980). Gesture and speech: Two aspects of the process of utterance. In Mary R. Key (Ed.), *Nonverbal communication and language* (pp. 207–227). The Hague: Mouton.
- Kendon, Adam (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, 23, 247–279.
- Kendon, Adam (2000). Language and gesture: Unity or duality. In David McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 47–63). Cambridge: Cambridge University Press.
- Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kimbara, Irene (2006). On gestural mimicry. *Gesture*, 6, 19–61.
- Kita, Sotaro (1993). Language and thought interface: A study of spontaneous gestures and Japanese mimetics. PhD thesis, University of Chicago, Chicago, IL.
- Kita, Sotaro (2000). How representational gestures help speaking. In David McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.
- Lakoff, George & Mark Johnson (2003). *Metaphors we live by*. Chicago: The University of Chicago Press.
- Lerner, Gene H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32 (2), 177–201.
- Linell, Per (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte: Information Age Publishing.
- Local, John (2005). On the interaction and phonetic design of collaborative completions. In William J. Hardcastle & Janet M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 263–282). Mahwah, NJ: Lawrence Erlbaum.

- Local, John & John Kelly (1986). Projection and “silences”: Notes on phonetic and conversational structure. *Human Studies*, 9, 185–204.
- Loehr, Daniel (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7, 179–214.
- Mandelbaum, Jenny (1989). Interpersonal activities in conversational storytelling. *Western Journal of Speech Communication*, 53 (2), 114–126.
- McNeill, David (1985). So you think gestures are nonverbal? *Psychological Review*, 92 (3), 350–371.
- McNeill, David (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.
- McNeill, David (2005). *Gesture and thought*. Chicago: The University of Chicago Press.
- Mondada, Lorenza (2006). Participants’ online analysis and multimodal practices: Projecting the end of the turn and the closing of the sequence. *Discourse Studies*, 8 (1), 117–129.
- Mondada, Lorenza (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9 (2), 195–226.
- Mondada, Lorenza (2011). Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, 43 (2), 542–552.
- Ogden, Richard (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetics Association*, 31, 139–152.
- Pomerantz, Anita (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In John M. Atkinson & John Heritage (Eds.), *Structures of social action* (pp. 57–101). Cambridge: Cambridge University Press.
- Raymond, Geoffrey (2003). Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review*, 68 (6), 939–967.
- Sacks, Harvey, Emanuel A. Schegloff, & Gail Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 (4), 696–735.
- Schegloff, Emanuel A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70 (6), 1075–1095.
- Schegloff, Emanuel A. (1984). On some gestures’ relation to talk. In J. Maxwell Atkinson & John Heritage (Eds.), *Structures of social action* (pp. 266–298). Cambridge: Cambridge University Press.
- Schegloff, Emanuel A. (1991). Conversation analysis and socially shared cognition. In Lauren B. Resnick, John M. Levine, & Stephanie D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 150–171). Washington: American Psychological Association.
- Schegloff, Emanuel A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97 (5), 1295–1345.
- Schegloff, Emanuel A. (1996). Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. In Barbara Fox (Ed.), *Studies in anaphora* (pp. 437–485). Amsterdam: John Benjamins.
- Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge: Cambridge University Press.
- Selting, Margret (2000). The construction of units in conversational talk. *Language in Society*, 29, 477–517.
- Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, & Jörg Bergmann (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion*, 10, 353–402. Retrieved from [www.gespraechsforschung-ozs.de](http://www.gespraechsforschung-ozs.de)
- Stivers, Tanya & Federico Rossano (2010). Mobilizing response. *Research on Language & Social Interaction*, 43 (1), 1–31.

- Streeck, Jürgen (2009). *Gesturecraft: The manu-facture of meaning*. Amsterdam: John Benjamins.
- Streeck, Jürgen & Ulrike Hartge (1992). Previews: Gestures at the transition place. In Peter Auer & Aldo Di Luzio (Eds.), *Contextualization of language* (pp.135–158). Amsterdam: John Benjamins.
- Svennevig, Jan (2008). Trying the easiest solution first in other-initiated repair. *Journal of Pragmatics*, 40 (2), 333–348.
- Walker, Gareth (2012). Coordination and interpretation of a vocal and visible resource: ‘Trail-off’ conjunctions. *Language & Speech*, 55 (1), 141–163.

## Appendix: Transcription conventions

### *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2) (Selting et al., 2009)*

#### Sequence structure

[ ]	
[ ]	Overlaps
=	“Latching”, a contribution starts immediately where a previous one ends

#### Breathing

°h/°hh/°hhh	In-breaths, 0.2–0.5 sec/0.5–0.8 sec/0.8–1.0 sec, respectively
h°/hh°/hhh°	Out-breaths, 0.2–0.5 sec/0.5–0.8 sec/0.8–1.0 sec, respectively,

#### Pauses

(.)	Micro-pause, below 0.2 sec
(-)/(--)/(---)	Pause, 0.2–0.5 sec/0.5–0.8 sec/0.8–1.0 sec, respectively
(1.0)	Longer pauses indicated by seconds

#### Durations

:/:/::	Prolongation of sound/syllable, 0.2–0.5 sec/0.5–0.8 sec/0.8–1.0 sec, respectively
--------	---

#### Accents

acCENT	Accented syllable in capital letters
ac'CENT	Rising pitch contour
ac`CENT	Falling pitch contour
ac˘CENT	Level pitch contour
ac^CENT	Falling-rising contour
acˆCENT	Rising-falling contour

#### Pitch movement at the end of intonation phrases

?	Rise to high
,	Rise to mid
-	Level
;	Fall to middle
.	Fall to low

#### Other conventions

ʔ	Glottal closure/hold
↑	Pitch step-up

↓	Pitch step-down
hu_uh	Two syllables
((head-move))	Non-verbal/visible productions or events
(he/you)	Candidate hearing/Possible candidates
<<p >>	Describing loudness and voice quality, e. g., p – piano, pp – pianissimo, f – forte

### *Gesture annotation (based on Kendon, 2004)*

~~~	Preparation of gesture stroke
***	Gesture stroke
***	Hold
-.-.-	Release of gesture
	Separating initiation/end of gesture unit, and gesture phrases

### *Authors' addresses*

Rein Ove Sikveland  
 Institute of Social and Cultural Anthropology  
 University of Oxford  
 51–53 Banbury Road  
 Oxford, OX2 6PE  
 UK

Richard Ogden  
 Department of Language and Linguistic  
 Science  
 University of York  
 Heslington, York, YO10 5DD  
 UK

rein.sikveland@anthro.ox.ac.uk

richard.ogden@york.ac.uk

### *About the authors*

**Rein Ove Sikveland**, post-doc researcher, is a phonetician whose work draws on conversation analysis, combined with linguistic and gesture analysis. He did his PhD in Linguistics at the University of York, entitled “Co-ordination of speech and gesture in time and sequence: Phonetic and visible detail in face-to-face interaction”.

**Richard Ogden**, senior lecturer, is a phonetician at the University of York who has worked on the phonetics of talk-in-interaction for over ten years. He has worked on a variety of languages and has a keen interest in sign languages.