Gesture in Modern South Arabian Languages: variation in multimodal constructions during task-based interaction[1]

Janet C.E. Watson & Jack Wilson

Abstract

This paper examines the role of gesture in communication in two Modern South Arabian languages, Mehri and Śḥerɛ̄t. We draw on audio-visual data of map and shape tasks collected in a purpose-built recording laboratory at the University of Leeds from three native speakers. This is the first paper to approach gesture in Modern South Arabian (MSAL), and the first to address the role of audio-visual data in the documentation of MSAL. In section 1, we address the need to appreciate gesture in communication, showing that while some texts can be understood from the written component alone, and some from the aural component alone, many orally delivered texts can only be fully understood from acknowledgement of the gesture component. We then discuss the language- and environment-specific aspect of gesture. In section 2, we discuss the co-expressiveness of gesture and speech. In section 3, we introduce the map and shape tasks, the equipment used, recording conditions and room layout, and the audio-visual data examined. In section 4, we present a semantic analysis of the gesture data. In section 5, we discuss our findings and provide suggestions for future research.

1 The importance of original recordings

The vast majority of texts published in relation to Semitic languages have been published in written form only, without access to the original audio or audio-visual recordings. There are exceptions, however, where recordings produced in the field have been collected and stored in archives. Notably, Johnstone's MSAL materials have been archived at Durham University and SOAS, London, and the sound recordings and notes have been so carefully produced and stored that several researchers have been able to work on the transcriptions, translations, notes and sound recordings (e.g. Hofstede 1998; Stroomer 1999, 2004; Rubin 2010, 2014).

There are three principal reasons why transcriptions of sound recordings are inadequate without access to the original recordings: first, even where individual researchers adopt a single transcription system, each researcher imprints their own interpretation of the phonological system. In working with transcribed materials, it is essential researchers are aware of the rationale of the transcriber. Secondly, human error often enters the transcription: several instances where Johnstone misinterpreted a word or phrase have been noted by Rubin (2010, 2014) and Stroomer (1999, 2004) with reference to the original sound recordings. And thirdly, without considerable cumbersome annotation, script is often inadequate in rendering the intonational and stylistic nuances of oral communication.

In general, greetings and short question/answer exchanges can be understood solely from the written component: the following text exchange with a Mehri consultant is unambiguous in most circumstances:

ād riddōna l-yaman
still return.FUT.MS the-Yemen

Are (you) going back to Yemen?

xāf ādī lardēd
perhaps still.1S return.1S
I may go back

By contrast, due to the post-position of the negative marker, complex negative clauses in written form can be ambiguous, as in:

wkōh klaṯk hēh ḏa-hēt haglāk lā
why told.2MS to.3MS that-2MS cooked.2MS NEG
Why did you tell him that you didn't cook? / Why didn't you tell him that you cooked?

The sound recording would provide the requisite intonational information to disambiguate the clause above; however, in many cases even written and aural information is insufficient for comprehension. Most texts in which the speaker attempts to persuade, cajole or force the listener to do something, and in which the message is external to the speech performance cannot be fully understood without considering gesture. This is particularly the case in texts involving a number of deictics. In the clause, *lḥamk tāmōl bi-ḏōmah ūṭōmah* 'I want you m.s. to do that like this', the aural and written components are insufficient to unpack the reference of *ḏōmah* 'this m.' and *ūṭōmah* 'like that'. The difficulty in interpreting deictics without the visual element led the first author to collect an initial body of audio-visual material of Mehri and Śḥerēt speakers in the field.

Our observation of face-to-face communication shows that gesture is almost invariably present, that it collaborates with the aural component to deliver meaning, and that it shows language- and environment-specificity. Different languages have access to a range of gestural/oral ways to express affirmation, for example, including: alveolar click accompanied by sharp raise of chin in dialects of Yemeni Arabic, hum and nod in English, sharp ingression of air in pronunciation of *ja* 'yes' in Norwegian. Different gesture/sound resources for expressing affirmation in different discourse contexts, for example: 'I follow you, continue', 'I would like to add something', 'I doubt what you say'. In direction texts, English speakers tend to use a pointed index finger to indicate straight on; in languages with a strong left-hand taboo, left will be indicated through the right hand moving across the body (Kita & Essegbey, 2001; Kita, 2009); in audio-visually recorded direction texts for Mehri and Śḥerēt recorded in the phonetics laboratory and in the field, speakers tend to use an expansive two-handed point for straight on, and wide full hand movements to indicate right or left. Finally, gesture is closely bound to the environment: in expressing the direction from which rain is coming, MSAL speakers use a hooked index finger: the direction of the finger indicates where the rain is coming from, and the hooked finger indicates the direction in which the rain is coming. The hooked finger iconically hooks and attracts the rain rather than pushing it away.

2 How are gesture and speech co-expressive?

This paper serves as a basic introduction to the importance of examining gesture in the documentation of MSAL. As a result, this section is brief and we direct those interested to various overviews of the area. Abner, Cooperrider, and Goldin-Meadow (2015) provide an excellent review for linguists. For a more detailed overview to theories related to gesture and its history, see Kendon (2004) or, for a complementary view, McNeill (2015). In this section, we describe gesture from a semantic and temporal perspective.

There are many ways in which gestures can be related to speech. However, here we focus on gestures that depict space. Just as linguists can analyse the content of speech in relation to the semantic content

of an utterance, gestures, too, can be analysed in terms of their semantics. In example 1 (taken from Kita and Özyürek (2003)), the same scene is being described in three languages. For English, the verb phrase "rolls down" encodes both the manner and the direction, whereas in Japanese and Turkish the direction is realised as a separate unit.

(1) English: *He rolls down the hill*
     Japanese: [korogat-te]   [saka-o   kudaru]
                 roll-CONNECTIVE slope-ACCUSATIVE   descend:PRESENT
                 (s/he) descends the slope, as (s/he) rolls.
     Turkish: [yuvarlan-arak]   [cadde-den   iniyor]
                 roll-CONNECTIVE street-ABLATIVE   descend:PRESENT
                 (s/he) descends on the street, as (s/he) rolls.

Kita and Özyürek (2003) demonstrated that speakers of Japanese and Turkish gestured, they were more likely to separate out the manner (i.e., rolling) from the direction (i.e., downward) in both speech and gesture. English speakers, on the other hand, were more likely to produce a gesture in which the manner and direction of the gesture was conflated. This data therefore, suggest that the way in which languages package semantic information grammatically also has an effect on the way language users gesture. This view has been referred to as the *interface hypothesis* (Kita and Özyürek, 2003) or *information packaging hypothesis* (Kita, 2000) and can be viewed in terms of Slobin's (1987) *thinking for speaking* model which suggests that information is packaged for speaking in relation to linguistic structure. Gesture reveals that such processes are not only relevant to linguistic practice, but are multimodal. Linguistic structure affect gesture segmentation. Below we will make an argument that the semantic structure of Mehri and Śḥerẽt has an effect on the gestures produced.

If speech and gesture are semantically linked, it is crucial to explain how it is possible to determine which gesture belongs to which linguistic utterance. Typically, co-expressive structure of gesture and speech is carefully timed to form a single temporally bounded gesture *unit* (Kendon, 2004). The gesture unit is the totality of visual activity bookended by two *rests* or *home* positions (Sacks & Schegloff, 2002). Home positions are moments of relaxation during which the articulators are not being employed productively. The gesture unit represents the entire movement from home and back again, and comprises three to five gesture *phases* (Kendon, 2004: 113–124). These phases are:

1. The preparation phase, consisting of the incipient stages of the gesture. This represents the initial movement away from home.
2. The (optional) pre-stroke hold in which the articulating hand/s is/are held in anticipation for the stroke (cf. McNeill, 2015: 6).
3. The stroke, which may be thought of as the nucleus of the gesture and is typically the most meaningful phase.
4. The (optional) post-stroke hold, during which the hand is held in its final position. This was originally observed by Kita (1993) and it allows the gesture producer to elongate the gesture's composition, often fitting it to the spoken elements of the utterance.
5. Lastly, the recovery phase is the movement back to home.

Gesture *phrases* are gesture units minus recovery and as such gesture units may comprise several gesture phrases, which each comprise several gesture phases. Kendon argues that gesture phrases closely collocate with the tone units (cf. Crystal & Davy, 1969: 24-40) of the accompanying speech. He further suggests that:

Tone units are packages of speech production identified by prosodic features which correspond to units of discourse meaning. In the same way, gesture phrases are units of visible bodily action identified by kinesic features which correspond to meaningful units of action such as pointing, a depiction, a pantomime or the enactment of a conventionalized gesture. (Kendon, 2004: 108)

3 Tasks and data analysis

Two tasks were used to elicit the data for this paper: the shape task and the map task. Both tasks were recorded in a purpose-built recording studio using a Canon XA20 video recorder and Audio-Technica AT2020 microphones. Both tasks involved two participants: the language consultant and the confederate played by the first author. The language consultant acted as Information Giver (IG) and the confederate acted as Information Follower (IF). Both participants were audio recorded, but only the language consultant was video recorded. Each participant had a workspace which appears on a wooden easel that occludes it from the other participant's view. The physical set up of the room is shown in fig. 1.
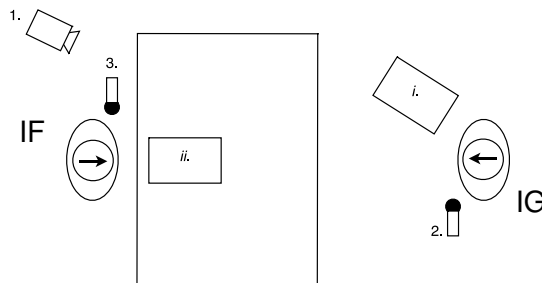

Figure 1. Physical set up, including: (1.) Camera, (2.) IG's microphone, (3.) IF's microphone, (i.) IG's workspace, (ii.) IF's workspace.

The shape task

For the shape task, IG was presented with a fixed order of two-dimensional shapes (see fig. 2) in a workspace in front of them. IF had the same shapes as IG but in IF's case they appeared on small card chips that can be moved around IF's workspace. The ultimate goal of the task is for IG to describe the shapes to IF so that IF can place their shapes in the same order.

Figure 2. IG's shapes

The map task

For the map task (cf. Anderson et al., 1991; Anderson, 2006), both IG and IF have a two-dimensional map (see fig. 3). IG's map contains several landmarks, start and finish points, and a route connecting the start and finish points. IF's map has the landmarks and start point, but lacks a finish point or route. The task is to physically recreate (i.e., draw), as closely as possible, IG's route on IF's map. In order to overcome the coordination problem created by the map task, the two participants must rely on a fractured and incomplete shared visual environment during communication.

The result of the manipulations in both tasks is that participants appear to superimpose their immediate visual environment into gesture space. The gesture space in front of them represents the map, and movements within it represent movements on the map. This plays an important semiotic role transforming, for example, a gesture which points to the space directly in front of a participant into gesture pointing at the map; as the index finger moves, it traces a line along the map. Such gestures have been referred to as tracing (Enfield, 2009) or drawing (Streeck, 2009).
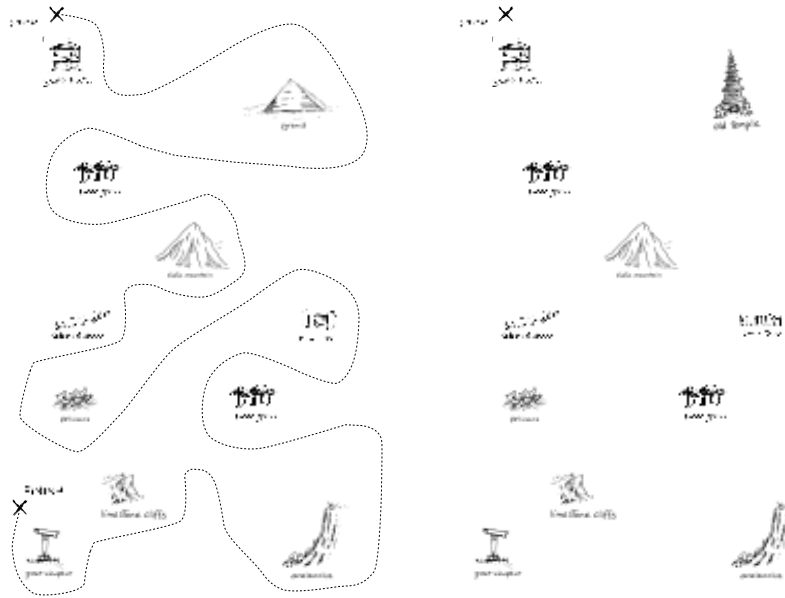

Figure 3. IG's map (left) and IF's map.

The speakers: background

Three male speakers of MSAL were recruited for these tasks. Many MSAL speakers are bilingual in two of the MSAL due to contact and the practice of marrying across language groups. In order to examine gesture in the two languages, we selected speakers with first-language knowledge of only one of the languages. All speakers are from Dhofar. Two, J043 and J108, are speakers of Śḥerēt in their early thirties and residents of the fishing village, Sadah. One, M001, is a 22-year-old speaker of Mehri from the gravel desert village of Rabkut. All speakers have been educated to secondary level; they have formal knowledge of Arabic, but use the local language in their daily lives. Note that the Mehri and Śḥerēt data analysed here is a sub-set of a >500-minute body of audio-visual data recorded in the field and in the laboratory from 17 male and 2 female speakers across the full educational range aged between 19 and mid-70s.

4. Analysis

The analysis presented here focusses on the semantic information conveyed through speech and gesture and the different ways in which such information is distributed across the two modalities. Our analysis draws, in part, on Talmy's (2000) semantics of space and motion events. Our aim is use semantic notions to qualitatively describe the representation of space in speech and gesture. The two tasks involve different depictions of space. The shape task involves a description of the *size* and *shape* of the referents described, and the map task involves constructions typically involving complex semantic relations which anchor the route described to landmarks within the map. In doing so,

participants typically describe the *orientation*, *direction*, and *manner* of the route relative to the landmarks.

The following analysis has been divided into sections relating to shapes in the shape task and particular parts of the map in the map task. In the following extracts, gesture phrases (see section 2 above) are delineated on the transcription line using square brackets with subscript numbers referring to the images in the following figures: for example, in Extract 1, [^Adaragah^A]$_1$ relates to fig. 4.1, and [^Atisʕīn^A]$_2$ relates to fig. 4.2. Whenever a hold is present, the concurrent words appear in **bold**. There is a degree of code-switching in the tasks, and Arabic terms are highlighted by raised ^A circumscript. Parenthesis containing numbers refer to pauses in seconds and milliseconds. Double parenthesis "(())" represent transcription comments. Omissions are marked as "((...))". Additionally, IF's interventions that do not coincide with a gesture from IG are omitted (indicated by ((IF…))). Word-for-word glosses are provided for complex utterances in plain type below the transcription line. English translations are provided below the word-for-word glosses in italics.[2]

4.1 Shape Task

4.1.1 Right-angled Triangle

The right-angled triangle pairs with the isosceles triangle because both require some qualifying modifier in order to distinguish one from the other. As we will see, the participants adopt different strategies when referring to this triangle.

In his description, the Mehri speaker, M001, focusses on the triangle's right angle. In five separate utterances, M001 specifically highlights the right-angled nature of the triangle. Accompanying each of his utterances is a gesture also highlighting the right angle. The establishment of the right angle gesturally affords M001 the possibility to refer to the hypotenuse through purely indexical means producing a *speech-framed* gesture (McNeill, 2009).

Extract 1

M001:   ^Aamṭallaṯ^A [^Adaragah^A]$_1$ [^Atisʕīn^A]$_2$ ((IF …)) [mġōran ūṭōmah]$_3$
          Triangle degree ninety ((IF…)) then like.that
          the triangle should be at ninety degrees ((IF …)) *then like that*

Figure 4.1–4.3

Extract 1 highlights M001's decomposition of the right-angled triangle into its three sides. The speech highlights the 90-degree corner of the triangle as a single element, whereas the gesture decomposes it into the two sides that intersect at the corner. Taken together, the speech and gesture collaboratively describe two thirds of the triangle. An important element of this extract relates to the temporal alignment of the gesture and the speech. The two words *daragah* 'degree' and *tisʕīn* 'ninety' are each accompanied by a gesture stroke. The first stroke depicts the vertical side of the triangle whereas the second depicts the horizontal side. There is no obvious semantic connection between either stroke and its temporally aligned lexical item or *lexical affiliate* (Schegloff, 1984); it is only when both lexical items and both strokes are taken together that the utterance works as a speech and gesture composite. Once established visually, the vertical side of the triangle acts as an anchor for the speech-framed gesture depicting the hypotenuse. These two-handed gesture sequences operate according to what Enfield (2009: 114–147) has analysed as a *symmetry-dominance* construction, displaying *gestural layering* (McNeill, 2005: 178). Such two-handed gestures are interesting because the non-dominant hand in the construction is presenting something that is not represented in the concurrent speech.

Extract 2

IF: wa-ḏōmah ykūn (1.5) śaymal aw ḥaymal (0.8) ḏōmah ḏo.. ykūn ūṭōmah aw ykūn ūṭōmah
        and-this.MS is.3MS left or right this.MS th.. is.3MS like.that or is.3MS like.that
        and this should be (1.5) to the left or the right? (0.8) this should be like this or like that?

M001:   lā [ḏōmah ykūn]₁ ḏə- [ḏōmah]₂ ((IF…)) [ḥaydiš ḥaymal tkūn bawmah]₂ [twōlī] [twōlī ḥaymal]₃
        no this.MS is th.. this.MS hand.2FS is.3FS here towards towards right
        no, this should be, this ((IF…)) to your right, it should be here, towards, towards the right

Figure 5.1–5.3

Following the turns depicted in Extract 1, IF questions M001 on whether the hypotenuse is to the right or left of the right angle; M001 once again responds deictically, holding his left forearm so as to form the vertical side of the triangle while producing a series of speech-framed gestures that first use the hand and forearm to represent the hypotenuse before bouncing his right arm to emphasise the fact that the hypotenuse is on the right-hand side of the triangle. This is emphasised and clarified in the speech and gesture in M001's next utterance where he first bounces his right arm in a manner similar to that of his first turn before sweeping his forearm outwards in a rightward direction. The first gesture coincides with *ḥaydiš ḥaymal* 'your right hand' and the second with *twōlī twōlī ḥaymal* 'towards, towards the right'. The repetition of *twōlī* seems to be a repair. This repair is mirrored in the gesture where M001 begins the outward path of his sweeping right forearm at the same time as the word *twōlī*, but restarts the movement as he restarts the word.

One of the Śḥerēt speakers, J108, adopts a completely different strategy in distinguishing the right-angled triangle. He introduces the right-angled triangle just with the Arabic word *amṯallaṯ* 'triangle' prompting IF to ask which triangle he is referring to. Extract 3 begins with J108's answer.

Extract 3

J108:   [ṭad ɛ-ḏ-īblɛ ibḳaʕš ṭanuh yikin]₁ [erɛšəš ḏ-īblɛ ṭaʰn]₂ (0.5) [ol ᴬsīdaᴬ lo fhaməš] ((IF …))
         one that-PART-leans.3MS put.2FS.3MS like.that is.3MS DEF.head.3MS PART-leans.3MS
         like.that NEG straight NEG understood.2FS
         one that leans, put it like that, its head leans like that (0.5) not straight, do you see? ((IF …))

J108:   [yikīn erɛš ɛ-ḏanun ᴬmṭallaṭᴬ]₃ ((IF …)) [ḏ-īblɛ ēmtə ᴬyasārᴬ]₃ (0.1) [ēmt ɛ-īdeš śemələt]₃
         is.3MS DEF.head of-this triangle ((IF …)) PART-leans.3MS towards left towards PART-
         hand.2FS left.FS
         the head of that triangle ((IF …)) leans towards the left (0.1) towards your left hand

Figure 6.1–6.3

In the beginning of J108's first turn as he produces the word *ɛ-ḏ-īblɛ* he uses his right hand to model the hypotenuse (with his index finger) and the horizontal side (with his thumb) directing attention to the gesture with the expression *ibḳaʕš ṭanuh* 'put it like that'. He then states *erɛšəš ḏ-īblɛ ṭaʰn* 'its head leans like that'. During the second part of the turn, he produces a speech-framed gesture with his left hand demonstrating the angle of the hypotenuse. He then repeats this utterance across two turns, first referring to the head of the triangle (*erɛš ɛ-ḏanun mṭallaṭ*) and pointing towards his workspace, then reproducing the same left-hand gesture depicting the direction in which the triangle is 'leaning'. There is an important addition to the utterance of *ḏ-īblɛ ēmtə yasār* in the form of a head tilt in a leftward direction. This head tilt is important because typically the verb *ḏ-īblɛ* describes animate objects. We argue that J108 is describing the triangle as if it is animate, using words like *erɛš* and *ḏ-īblɛ*, and allowing him to embody the triangle in his bodily movement as well as his manual gesture. This argument finds support later in the interaction when J108 is distinguishing the isosceles from the right-angled triangle.

Extract 4

J108:   [o lebre ḏek lo]₁ [d-iskof də-šiśḳī lo]₂
         NEG like that.MS NEG that-sits.3MS PART-lies.on.side.3MS NEG
         not like that one that sits lying on its side

Figure 7.1–7.2

While producing the utterance in Extract 4, J108 ostensively leans over to his left-hand side holding his hands in front of him as depictions of the two non-horizontal sides of the triangle.

8

4.1.2 Diamond

Our diamond is actually a square rotated 45 degrees. Therefore, as with the right-angled triangle, the participants need to elaborate in order to distinguish it from the other square. During his description of the diamond, M001 repeatedly touches the tips of his fingers and the heels of his hands together, forming the top and bottom angles of the diamond. However, rather than describing the diamond, he repeatedly produces the word *ūṭōmah* 'like this' framing his gesture rather than referring to the diamond directly.

J108's strategy for referring to the diamond is quite different both linguistically and gesturally, and here we present transcription and figures.

Extract 5

J108:   āxərī [^Amrabbaʕ^A nīṣàn]₁ [ibḳaʕš l-erɛš]₂ (0.3) [o tibḳaʕš ^Asīda^A lo]₃ (0.5) [ibḳaʕš yol e-rɛš]₄
         last.one square small put.3MS to-DEF.head NEG put.3MS straight NEG put.3MS to DEF.head
         The last one is a small square, put it on the head (0.3), don't put it straight (0.5), put it on the head

Figure 8.1–8.4

J108 describes the diamond as the small square instructing IF to put it on its head. This is distinct from M001's strategy because it focusses on the placement of the diamond in relation to some assumed flat plane at its base, which acts as the ground for the spatial construction. This is directly reflected in his gesture which represents the diamond; he does not use a gesture to trace the entire shape, but models its lowest point with his index finger pointing downwards. His left hand forms the ground upon which the point can be placed. In fig. 8.3, J108 creates a counterfactual situation in which the small square is straight, co-expressive with a gesture that reflects this. In the gesture, J108 depicts what it would look like if the shape was straight using the same left-hand shape to represent the flat ground, only this time using a parallel right hand to represent the shape: a shape with its side parallel to the horizontal axis is 'straight' and one with its side at 45 degrees to horizontal axis is not. Finally, in fig. 8.4, J108 reiterates his first gesture and speech rearticulating the orientation in which the shape should be placed. The diamond is mentioned once again later in the interaction.

Extract 6

J108:   ^Aāxar šī mrabbaʕ^A nīṣan ^Alākin^A ol tibḳaš [^Asīda^A lo]₁ ((IF…)) [ḳalaʕš yėṣer l-erɛš]₂
         last thing square small but NEG put.3MS straight NEG ((IF…)) put.3MS stand.3MS to-
         DEF.head
         the last thing, the small square, but don't put it straight ((IF …)) let it stand on the head

IF: [ḥayš̃ōf]₃
         okay

J108:   [l-^Aezāwiya^A fhaməš̃]₄
         to-DEF.corner understood.2FS
         on the corner, you see?

Figure 9.1–9.4

In his first turn in fig. 9, we see an identical command to the one produced earlier, namely *ol tibḳaš* ᴬ*sīda*ᴬ *lo*. This is once again accompanied by a similar gesture (fig. 9.1); however this time it involves a different orientation of the hands. This time his hands are laid prone on top of each other. J108's second turn is similar to the one produced earlier in the interaction; however here he produces two gestures: the first (fig. 9.2) is almost identical to the one described above for the right-angled triangle (fig. 7) where J108 embodies the shape holding his hands out in front of him to represent prominent sides and leaning in the direction in which the shape is oriented; the second (fig. 9.3), which occurs after his utterance, represents the shape using the thumb and index finger of the right hand and rotating the hand 45 degrees. This gesture seems to depict the two states in his previous two turns (i.e., first straight, then not straight). The puzzle here then is the second gesture which seems out of place and more appropriately depicts a shape like the right-angled triangle. We believe the reason for this is down to the verb *yeṣer* 'it stands', which like *ḏ-īblɛ* is typically used to describe animate objects. If the animacy of this term triggers embodiment, then the leaning is a manifestation of the non-straightness that J108 has been discussing, rather than iconically representing the shape. In his last turn (fig. 9.4), J108 highlights the corner (*l-ezāwiya*) and produces a gesture similar to the one produced by M001, with both hands touching at the heels to represent the bottom half of the diamond.

4.1.3 Pentagon

J043 produced an interesting sequence of utterances in discussing the pentagon. The extracts presented here are taken from an extended discussion of the shape.

Extract 7

IF: wa-tkīn (0.9) tkīn ṭaʰn
        CONJ-is.3FS is.3FS like.that
        and it should be (0.9) it should be like that?

J043:   [ᴬaywahᴬ ṭaʰn]₁
        yes, like that

IF: min ṭaʰn
        or like that

J043:   lā (( )) baʕlit xĩš ᴬzwāyaᴬ
        No object.with.FS five.FS corners
        no ((J043 produces gesture depicted in fig. 10.2–10.4)) *that with five sides*

Figure 10.1–10.4

Up to this point in the task, J043 has referred to the pentagon as the shape with five corners. IF requests clarification regarding the shape. Because we only filmed the language consultants we do not know what IF is describing. However, it is clear that IF first asks a question presenting two alternatives, the first J043 agrees with, the second he does not. Concurrent with his utterance of *aywah ṭaʰn*, J043 begins to depict the top two sides of the pentagon with flat, prone hands to represent the sides touching his fingertips together to represent the angle (fig. 10.1). However, when IF utters *min ṭaʰn*, presenting the second alternative, J043's dissent triggers a more elaborate gesture. In this gesture J043's left hand is held, depicting the top left side of the pentagon, but his right hand depicts each side, beginning with the top right and ending with the bottom left, thereby completing the whole shape. This gesture seems to arise as a result of J043's focus on the number of sides of the shape. Later on in this sequence, J043 changes his focus.

Extract 8

J043:   [min ḥaṯih ay min ḥaṯih ṭaʰn]₁
          from above yes from above like.that
          from above, yes, from above like that
IF: ay ay (0.5) ahah ḥayš̃ōf
          yes yes yes okay
          okay okay (0.5) yes, okay
J043:   ḥayš̃ōf ((IF…)) [nafs ɛ-tēg]₂
          Okay ((IF…)) same.as PART-crown
          good ((IF…)) like a crown

Figure 11.1–11.2

In this extract, J043 shifts his attention to the top of the shape. Co-occurring with the word *ḥaṯih* 'above', J043 repeats his earlier gesture (fig. 11.1) depicting the top two sides of the shape, a gesture he holds for the rest of his turn. Following this are a series of affirmative comments (not transcribed above) before he compares the shape to a crown. His concurrent gesture (fig. 11.2) involves two fingers pointing to the ceiling. If we took this utterance on its own, the gesture may not seem to be particularly relevant to the utterance. However, if we consider that the top of the shape is already discourse prominent then the meaning of this gesture changes. In pointing upwards, he is pointing towards the top of a crown. This gesture therefore ties J043's previous utterance to the current one. The shape is like a crown because it has a pointed top.

4.2 Map Task

In the following analysis, we focus on the utterances describing direction and orientation that accompanied the description of the route between the Pelicans and the Broken Gate (see figure 3). This section was chosen because it is representative of the larger corpus and also because it represents

11

the largest single trajectory on the map. As we will show, the two participants we focus on use a variety of multimodal strategies to describe this section of the map.

Here we compare M001's description of the route between the pelicans and broken gate (Extract 9) with J108's description of the same route (Extract 10).

Extract 9

M001: [ḳəfēdī xōṭar]₁ ((IF…)) [hāl min śərūḵ]₂ [ḏ-āḵāb]₂ [l-śərūḵ]₂[əsan]₂ [śaymal]₂ ((IF…))
go.down.2FS down ((IF…)) to from side PART-birds to-side.3FPL left ((IF…))
go down ((IF…)) by, to the side of the birds, to their left side ((IF…))

M001: [wa-nkay m-nxalī āḵāb]₃ ((IF…)) [wa-śbēbī aġa:::wf ta-tśbēbī hāl hāl]₄ [agīdōr aṯ-ṯībar]
CONJ-come.2FS from-under birds ((IF…)) CONJ-go.up.2FS up until-go.up.2FS to to DEF.gate
part-broke.3MS
and come from under the birds ((IF…)) and go up and up until you go up to the broken fence

IF: ḏ-aṯīrōb ahah
PART-sticks yes
of sticks, okay

M001: [ahah śbēbī ṭayrah aġawf]₅ ((IF…)) [wa-ḳfēdī min śərūḵəh]₅ [ḍa-ḥḥaymal]₆
yes go.up.2FS above.3MS up ((IF…)) CONJ-come.down.2FS from side.3MS part-DEF.right
yes, go up over it ((IF…)) and come down on the right side of it

Figure 12.1–12.6

In Extract 9, M001 first describes the route as it travels down the left of the pelicans, using *ḳəfēdī* ('go down') which specifies the downward direction of the route but modifying it with the adverb *xōṭar* ('downwards') further specifying direction of travel. His gesture (fig. 12.1) at this point is co-expressive with the verb, with the adverb realised during a post-stroke hold. His next turn (fig. 12.2) specifies the leftward orientation of the route to the landmark. M001 depicts this leftward orientation with his left hand, holding his right hand in the same position as it was at the end of the stroke depicting downwards. Depiction of the leftward orientation is not produced as a leftward path with a single stroke, but as a leftward beat on every other syllable. This timing is not coincidental but brings focus to the fact that the whole utterance is about this leftward orientation. Although there is only space to show a few examples in this analysis, we believe that this example highlights an interesting characteristic of gestures accompanying Mehri and Śḥerēt. In every example we have analysed

(including those not presented here), speakers produce separate direction and orientation gestures. Equally, Mehri seems to unpack spoken spatial description so that route direction and route orientation appear as separate linguistic units. However, this fact seems to be limited sagittal plane when speakers are describing the route as being to the left or the right of a landmark. This suggests that direction and orientation represent distinct conceptual units for M001. In Extract 9, the right-hand hold gesture accompanied by the left-hand beat gesture (figure 12.2) is a visual representation of the linguistic disconnection between direction and orientation. This is in line with the information packaging hypothesis (Kita, 2000) where gesture structure is package relative to the process of *thinking for speaking* (Slobin, 1987). In another example that we do not have space to show here, M001 is shown to produce identical gestures when describing the orientation and the direction of a different section of the route.

Next, M001 describes the route as it travels under the pelicans before describing its upward trajectory towards the broken gate. He produces a gesture depicting the route throughout both of these utterances, pausing at the end of his first turn and making eye contact with IF in order to elicit confirmation that she has understood. Interestingly, M001 describes the route as *wa-nkay m-nxalī āḵāb* ('and come from under the birds') where 'under' is post-modifying 'come'. In this sense the direction and the orientation form a single linguistic unit, because direction is implied by the fact that M001's description begins with the routes origin (i.e., 'under the birds') and assumes its goal. What's more, M001 does not produce a separate 'under' gesture, which further emphasises the idea that sagittal orientation is realised as a separate unit. Another interesting feature with the second turn is the acoustic emphasis placed on *aġa:::wf* through rising pitch and elongation of the vowel. The route between pelicans and broken gate represents the longest straight line on the entire map. Therefore, it seems that M001 is depicting the length and upward trajectory of the route both through gestural and acoustic means. This example emphasises that multimodality is not simply tied to different channels of expression (e.g., speech and gesture) but can exist within and across channels (cf. Goodwin, 2009).

In M001's last turn he describes the route as it comes down the right-hand side of the broken gate. Here, we once again see the direction and orientation realised as separate gestural elements. First, a gesture depicting the downward trajectory is co-expressed with *wa-ḵfēdī min śərūḵəh* (lit. 'and come down from the side of it') and then a gesture comprising rightward beats, similar to the one represented in fig. 14.2, is produced. This gesture also has a rightward stroke on every other syllable. Second a gesture This time, however, the orientation gesture is produced with the same hand as the hand used to depict direction.

Extract 10
J108:   [tirfaʕ ḥaṭih l-īdēš]₅ ɛməlēt ((IF…)) [irfiʕ irfiʕ irfiʕ irfi:ʕ]₆ bi-ḥaṣɛr jurɛn bi-ᴬbawwābahᴬ dịhⁿ śīnəšīeᴬbawwābahᴬ dᶦinuh ((IF…))
         go.up.2FS up to-hand.2FS right.FS ((IF…)) go.up.2FS go.up.2FS go.up.2FS go.up.2FS CONJ-when pass.1PL by-gate this.MS saw.2FS DEF.gate this.MS
         go up to your right ((IF…)) go up, up, up, up, and as we passed by that gate, can you see that gate? ((IF…))
J108:   [irfiʕ ḥaṭih]₇ ((IF…)) [irfiʕ ḥaṭi:h]₈ (0.3) [l-īdēš ṭaʰn ykin ᴬmusta..ᴬ]₉ [mɛyil] [ṭan l-īdēš]₁₀ [ɛməlēt **xaṭṭ**]₁₁ ((IF…)) [l-īdēš ɛməl...]₁₂
         Go.up.2FS up ((IF…)) go.up.2FS up to-hand.2FS like.that is.3MS straig.. leaning like.that to-hand.2FS right.FS **line** to-hand.2FS righ..
         go up [IF …] go up (0.3) to your hand like that, it is straig.., inclining like that, to your right hand, the path ((IF…)) to your right hand
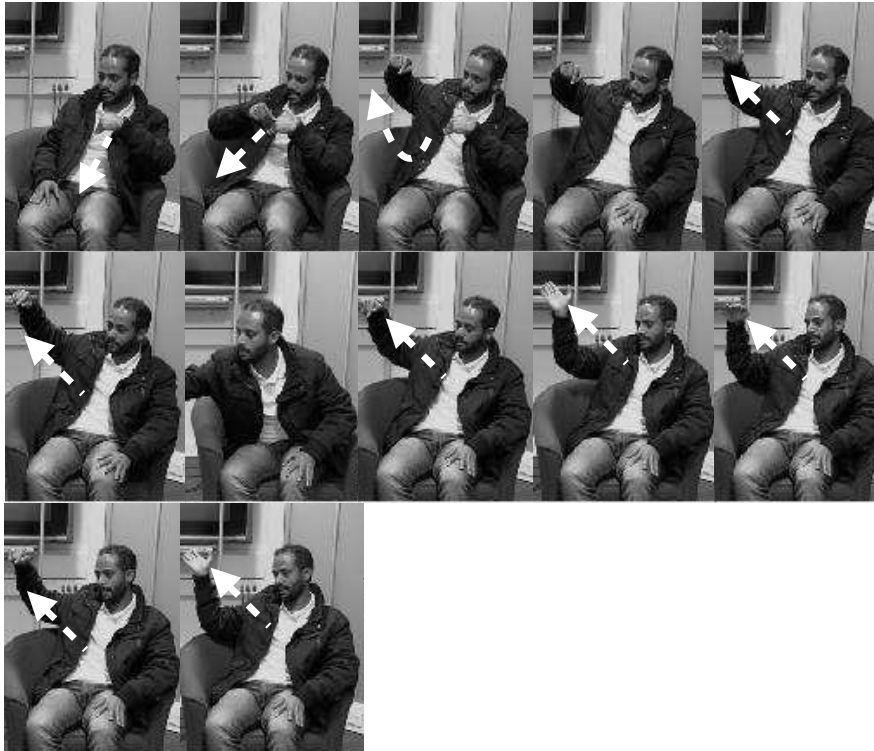
13

Figure 13.1–13.12

In extract 10, J108 uses speech to emphasise the length of the portion of route being described, as seen for Extract 9; however, whereas M001 raises pitch and elongates the vowel, J108 reduplicates the word: *irfiʕ irfiʕ irfiʕ irfi:ʕ*, elongating the final vowel. This turn is co-expressive with a stroke (fig. 13.6) depicting the route shape. J108's next turn repeats this idea of going up, however, the second time he says *irfiʕ ḥaṭi:h* (lit. 'go up upwards', fig. 13.8) he raises pitch and elongates the final vowel in a manner identical to M001. Following this, J108 describes the direction the line is moving in a number of different ways, each occurring with a different gesture. First, J108 produces a gesture (fig. 13.9) co-expressive with *l-īdēŝ ṭaʰn ykin musta...* Here the diagonal direction of the route is expressed through the speech-framed gesture with the manner of the route expressed linguistically through the incomplete Arabic word *musta....* Second, J108 produces an entire gesture phrase with the word *mēyil* (reproducing his embodied spatial description in extract 4). This gesture (fig. 13.10) is repeated with *ṭan l-īdēŝ* before finally being replaced with a pointing hand for the final part of the turn *ɛmǝlēt xaṭṭ* with the stroke of the gesture (fig. 13.11) falling on the word *xaṭṭ*. In this utterance, J108 is describing the direction using the relative orientation of the route to the observer. Interestingly, like the orientation terms shown in extract 9 above, orientation is always described (and gestured) separately from direction. Furthermore, extract 10 shows that this seems to be tied to the sagittal plane.

5. Discussion

The above analysis demonstrates that there is a richness to gesture that can be lost when visual data is not captured. Through this analysis we have focussed on three speakers. The extracts show a range of multimodal strategies, demonstrating individual tendencies to use speech and gesture in particular ways. One thing that unites all speakers is that the speech and gesture are co-expressive; they are not two isolated channels through which speakers communicate. It is only when speech and gesture are taken together that the fullness of their expressions can be grasped.

Furthermore, the co-expressive nature of the utterances described above highlight another crucial (and potentially unexplored feature of Mehri and Śḥerɛ̄t). Within the gesture literature, it is widely acknowledged that the packaging of information in gesture reflects the syntactic structure of the language used. We have shown that this is also the case for Mehri and Śḥerɛ̄t. One finding that has emerged in these analyses relates to the use of right and left orientation terms in both languages. We have shown that orientation and direction appear as separate conceptual elements that are realised as separate gestures. We do not find direction being expressed linguistically as part of the orientation of the route to the landmark. In our lab data, and impressionistically in our field data, direction and orientation always appear as separate speech phrases or turns, e.g.: 'go down from it[direction], on your right hand[orientation]'. Further, we never see a gesture stroke connect the direction and the orientation; the two are always realised as separate gesture phrases. In the context of gesture research which has focussed on manner and path, this finding is novel because it highlights the intrinsic relationship between orientation and direction. What this finding demonstrates is that an understanding of the segmentation of gesture can provide insights into the structure of language. In this paper we are arguing that there is a linguistic separation of direction and orientation in both Mehri and Śḥerɛ̄t. Ultimately, this paper shows that gesture is not just crucial for expression, but can provide insights into linguistic features which were once hidden to researchers. Future work will further explore the insights gesture can offer in the analysis of Mehri and Śḥerɛ̄t.

References

Abner, N., Cooperrider, K., and Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass*, 9(11):437–449.

Anderson, A.H. 2006. Achieving understanding in face-to-face and video-mediated multiparty interactions. *Discourse Processes*, 41(3): 251–287.

Anderson, A.H., et al. 1991. The HCRC map task corpus. *Language and Speech*, 34(4): 351–366.

Clark, H.H. 1996. *Using language*. CUP.

Crystal, D. & Davy, D. 1969. *Investigating English Style*. Longman: New York.

Enfield, N.J. 2009. *The anatomy of meaning: Speech, gesture, and composite utterances*. CUP.

Goodwin, C. 2013. The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46(1): 8–23.

Hofstede, A. 1998. Syntax of Jibbāli. PhD thesis: University of Manchester.

Johnstone, T.M. 1987. *Mehri Lexicon and English–Mehri Word-List*. Routledge: London.

Kendon, A. 2004. *Gesture: Visible action as utterance*. CUP.

Kita, S. 1993. *Language and thought interface: A study of spontaneous gestures and Japanese mimetics*. PhD thesis, University of Chicago, IL.

Kita, S. (2000). How representational gestures help speaking. In McNeill, D., editor, *Language and gesture*, pp. 162–185. Cambridge University Press.

Kita, S. 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24(2):145–167.

Kita, S. & Essegbey, J. 2001. Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. *Gesture*, 1:73–94.

Kita, S. and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, 48(1):16–32.

McNeill, D. 2005. *Gesture and thought*. University of Chicago Press: London.

McNeill, D. 2009. Gesture and communication. In Mey, J. L., editor, *Concise Encyclopedia of Pragmatics*, pp. 299–307. Elsevier: Oxford.

McNeill, D. 2015. *Why we gesture: The surprising role of hand movements in communication*. CUP.

Rubin, A. 2010. *The Mehri language of Oman*. Brill: Leiden.

Rubin, A. 2014. *The Jibbāli (Shaḥri) language of Oman*. Brill: Leiden.

Sacks, H. & Schegloff, E.A. 2002. Home position. *Gesture*, 2(2): 133–146.

Schegloff, E.A. 1984. On some gestures' relation to talk. In Atkinson, J. & Heritage, E.J. (eds), *Structures of Social Action: Studies in Conversation Analysis*. 266–296. CUP.

Sima, A. 2009. *Mehri-Texte aus der jemenitischen Šarqiyyah*. Harrassowitz: Wiesbaden.

Slobin, D. I. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the 13th annual meeting of the Berkeley Linguistic Society meeting*, pp. 435–445).

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. John Benjamins Publishing: Amsterdam.

Stroomer, H. 1999. *Mehri texts from Oman: Based on the field materials of T.M. Johnstone*. Harrassowitz: Wiesbaden.

Stroomer, H. 2004. *Harsusi texts from Oman: Based on the field materials of T.M. Johnstone*. Harrassowitz: Wiesbaden.

Talmy, L. (2000). Toward a cognitive semantics, vol. I: Concept structuring systems. MIT Press, Cambridge, MA.

Watson, J.C.E. 2012. *The structure of Mehri*. Harrassowitz: Wiesbaden.