# On secret signals, naughty signs and public speech: The gestural foundations of communication

*Dafydd Gibbon*

Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany

gibbon@uni-bielefeld.de

**Abstract**

Why are gestures so fascinating – and so powerful? In some countries certain gestures can be very expensive if the gesturee feels unjustifiably insulted by the gesturer and takes him (perhaps more often than her) to court. Many gestures are strongly lexicalised, and few of these are universal. Some gestures are familiar to everyone in a given culture, such as the culture-specific lexical gestures for dialogue acts: greeting by waving or by different types of face, hand or lip kissing; applauding by clapping; surprise by eyebrow-raising; disgust by nose-wrinkling; agreement by nodding; disagreement by head-shaking. Some gestural functions, such as deixis, are universal, even if the actual means of gesturing differs from culture to culture. Some gestures involve more than one part of the body, such as applauding, or finger touching face (or specifically nose) for insulting. Some gestures involve body contact between gesturers, such as backslapping, and others may involve body contact and are (almost) necessarily accompanied by words, such as greeting by handshaking, accompanied by pragmatic idioms such as "How do you do?", or slapping right hands as in "Gimme five!". Some gestures are kept secret, either to the individual, as with fingers crossed behind the back to turn a lie into a 'white lie', or the special handshakes of secret societies. Some gestures are recognisably public rhetorical gestures. Some gestures generate sound, such as clapping, stamping, finger-snapping, air kissing. Some gestures are not categorised by gesturers as gestures at all, such as the gestures of the speech organs, which are transduced into sound. Some gestures are organised on a scale of increasing complexity, such as the gestures on the formality-intimacy scale: waving, hand-shaking, back-slapping, hugging, kissing, caressing, love-making.

There have been many attempts to categorise gestures and related categories of posture and spatial positioning, both autonomously and in broader multimodal communication contexts, by linguists, anthropologists, psychologists and sociologists, and most recently by video game and robotics specialists in computer science. In this contribution, a proposal will be made from a strictly linguistic point of view, based on research work on designing a formal gesture lexicon, towards developing an ontology which will make sense of all these examples of gesture and gesture-speech relations, including gesture grammar, gesture semantics and gesture pragmatics, and for applying such an ontology not only in creating re-usable gestural resources but also in the computational processing of gesture.

## 1 Introduction

### 1.1 A linguistic approach to gesture

Churchill's use of the victory sign, the Hitler salute, an air kiss to a parting close friend, a wave, beckoning with a finger, a dismissive hand movement, cupping a hand around the ear, the f-sign or 'the finger': these gestures are, at first glance, simple to categorise: they are all hand movements with some communicative function; they are all signs. But the details of the communicative functions, and the differences between the gestures need to be categorised, as well as the similarities, and this is perhaps not as simple as it may seem.

The present approach is to use a linguistic metaphor. That is, gestural communication is treated as spoken language, with the exception of a switch from the auditory to the visual modality. The forms, structures and functions of gestures will be taken to be analogous to the forms, structures and functions of speech. Specifically, the relation of gestures to speech will be taken to be analogous to the relation of prosody to speech. On this basis, a systematisation of gesture from paradigmatic (classificatory, taxxonomic) and syntagmatic (compositional, mereonomic) perspectives will be proposed, though not all details of the speech-gesture parallels can be detalt with in the present context.

There is a precedent for a linguistic approach in the 'sign language phonology' of communication by hearing-impaired communicators. Use of the analogy of spoken language will perhaps lead to artefactual constructs and distinctions; this remains to be seen. The linguistics of spoken language is certainly much more advanced than the study of gesture, so the linguistic approach may at least prove to be a useful heuristic. Indeed, there is also a precedent for the inverse analogy, 'speech is gesture', systematised in the paradigms of Articulatory Phonology (Browman & Goldstein 1992) and Time Map Phonology (Carson-Berndsen 1998).

The present approach contrasts with that of McNeill (1992), in which gesture is not morphemic, not compositional, not defined in terms of form-meaning conventions, but rather global, in that the meanings of the parts are determined by the meaning of the whole in the overall context in which they are embedded. An initial account of the present approach is given by Gibbon (2005).

## 1.2 A preliminary definition of gesture

From the perspective of the linguistics of discourse, all of the gestures mentioned at the start either initiate or terminate a phase of phatic interaction, either starting or ending some kind of communicative encounter or sub-encounter (i.e. episode in the encounter). As a first approximation to characterising the differences in the forms and functions of gestures, a standard strategy of *definitio per genera proxima et differentia specifica*, can be followed, and could run something like this (here concentrating only on manual gestures):

> A communicative gesture is a movement of the hands,
> (a) with one limb only, hand/arm/fingers shaped like *X*, with (no) contact with another part of the body, with no contact with an interlocutor, starting/finishing/positioned at *Y*, and
> (b) with positive or negative associations or sanctions, initiating/terminating a dialogue act *A* with a goal *B*, in a social configuration *C*.

Each *differens specificum* constitutes a position on a dimension in a multi-dimensional semiotic quality space within which the different gestures are located, and each *genus proximum* constitutes a region in this semiotic quality space. Of course anything approaching a full definition will require many more dimensions than this, and some of these will be discussed below.

A more formal explication of this kind of quality space could use attribute-value structures from formal syntax and semantics, or feature bundles from phonology; plus constraints restricting their internal structure and compositional operations for combining such structures into larger communicative acts. A formal representation is not only necessary for a consistent, complete, sound and precise account of gestural communication itself, but necessary also for the integration of gestural communication into an overall theory of multimodal communication with speech and gesture. And, for those interested in these domains, for application in forensics, robotics, and video games development. In he present context, pointers to more details will be given, but not a complete ontological model.

# 2 What gestures are – characterising ‘gesture’

## 2.1 Categorisation of gesture forms

The history of known gesture studies goes back to the study of rhetoric in ancient Greece and Rome, to ancient holy scriptures, and extends to the thespian lore of gesture in dramatic acting. An early interdisciplinary linguistically informed approach to speech and gesture analysis is Pittenger & al. (1960). More recent scientific approaches have rarely been from a linguistic point of view, but have mainly been related either to behavioural and social interaction studies, where the functions of gesture are focussed rather than their forms, or to gesture modelling in artificial intelligence and robotics, where the forms are more important than the functions, non-semiotic gestures such as grasping and turning are more important than semiotic gestures, and latter are often reduced to sets of pointing stereotypes for denoting positions in space. More recently the term “gesture” has come to be associated with computer input devices such as the mouse (arm and wrist movements directing the mouse in two dimensions, button clicking, wheel turning) and the touchpad (finger movements, generally with one finger, but more recently with two, in the future perhaps more than two), as well as joysticks and body movement detectors in Wii-type video games.

In an overview paper, Kendon (1996), one of the pioneers of modern gesture studies, discusses a range of previous approaches, and proposes an agenda for future studies. To some extent, this agenda is already being fulfilled, but studies are very fragmented and driven by a host of different interests, from the sign languages of the hearing-impaired to software development for robotic systems. The following excerpt from his 1996 paper (cited from the web version) nicely summarises the formal features of a gesture, which he refers to as a ‘phrase of action’: “This 'strand' of activity (which we also refer to when we use the term 'gesture' or 'gesticulation') has certain characteristics which distinguish it from other kinds of activity (such as practical actions, postural adjustments, orientation changes, self-manipulations, and so forth).” The characteristics referred by Kendon include:

1. Movement is away from a rest position and back to a rest position.
2. The movement has a ‘peak’, ‘centre’ or ‘stroke’.
3. The temporal boundaries of the movement are clearly demarcated (unlike gradual changes in orientation or posture).
4. The movement is symmetrical, contrast to many other actions. A reverse spooled video may be difficult to distinguish from the forward spooled version video.

The fundamentals of Kendon’s description have been adopted by many other researches, and is well-adapted to formalisation. The characterisation can also apply to speech: the gestures of the larynx (phonation types, glottal stop), the velum and uvula (nasality), the tongue and lips (vowel and consonant configurations). In phonetics, the ‘stroke’ is usually defined in phonetics in terms of a ‘target’ position (cf. Articulatory Phonology, Browman & Goldstein 199) which is aimed at, but not always be reached, depending on phonostylistic factors.

What is lacking in this characterisation is an account of the compositionality of communication: the combinatorics of gestures with other gestures, in particular with speech. It is tempting to extend the concept of ‘prosody’ to gesture, because in many cases gestures relate to speech as prosody relates to speech. Instead, the term *modality* is conventionally used to refer to these distinct kinds of communication:

> A *modality* is a communication channel characterised by a pair of human motor output and sensory input organs.

The term *multimodal* refers to a combination of modalities used in real-life communication, and the combinations of modalities which occur are *multimodal complexes*. A functionally distinct use of the same articulators and the same input organ, such as prosody in speech, is a *submodality*. It should be noted in passing that the terms ‘multimodal communication’ and ‘multimedia communication’ are often used interhangeably. However, technically speaking, multimodal communication covers human input/output relations, and multimedia communication covers the

different technical channels which may be used simultaneously, possibly with the same modality (e.g. speech and music; text, photographs and video recordings).

## 2.2 Categorisation of gesture functions

The well-known categorisation by McNeill (1992) is functional, with only very basic attention to the forms of gestures. The following list adds two further function categories (*emblems* and *affectives*) to McNeill's five main categories:

1. *Iconics*, where the gesture resembles the referent (e.g. describing an action or shape of an object with the hands).
2. *Metaphorics*, where the *vehicle* (the gesture) relates in one of a number of metaphorical ways to the *tenor* (non-literal meaning) of the gesture, e.g. indicating a container or conduit for ideas , or a gift of an idea or suggestion (cf. Lakoff & Johnson 1980).
3. *Beats*, where the hand, head, eybrows move roughly in synchrony with the rhythm of often emphatic speech, mark a sequence, or a hiatus such as a change of theme or focus.
4. *Cohesives*, which create a gestalt in gesture space which is coextensive with a spoken utterance or – hierarchically – with its parts.
5. *Deictics*, which may indicate an actual physical position, size, distance or direction, but may also place concepts metaphorically in physical gesture space
6. *Emblems*, which are fairly highly conventionalised, lexicalised gestures, and constitute the most well-known tye of gesture.
7. *Affectives*, which display emotional states and events.

The phatic greeting and farewell gestures noted at the beginning of this contribution belong to the category of emblems. Functionally, these phatic gestures are like interjections, in that they have a relatively fixed but often hard to define form-function relationship, and they do not fit into the regular flow of speech, but have an autonomous attention-getting, channel-creating or emotional status. The same applies to the chant-like stylised phatic intonation (Gibbon 1976) used in calling, routine lists and corrections, and with some interjection-like greetings ("Hello-o!") and farewells ("By-ye!"). Other gesture emblems are more clearly related to the main parts of speech of a language and, like other parts of speech, are highly language specific or culture specific: various configurations of hand and fingers with a wide range of clearly identificable meanings such as success, pleasure, idiocy, cuckoldry, disgust, eating, drinking and telephoning.

Note that McNeill's categories are not necessarily mutually exclusive types, but rather parameters whose values can co-occur in any given gesture: emblems can be iconic and metaphorical, for example, holding the hands wide apart to indicate the great importance of some issue.

# 3 What gestures are not – delimiting 'gesture'

## 3.1 Other gesture-like semiotic movements

The term "gesture" could be – and sometimes is – used to cover any kind of non-verbal (and indeed also verbal) body movement which can be given a semiotic interpretation. This general use of the term is not necessarily the most useful, however. The important point, though, is to clarify the concepts, and not worry too much about the terms except to define them and use them consistently. There are many studies of behaviours in a range of neighbouring sciences, but these cannot be considered in any detail here. A number of useful distinctions which belong in an overall ontology of communication will be proposed in the following discussion.

### 3.1.1 Practical behaviour

Gestures, especially iconic and metaphorical gestures resemble other kinds of bodily activity which have no semiotic import:

1. Fortuitously triggered or concomitant movements such as stumbling, stubbing the toe, banging the head, staggering.
2. Deliberate but practised and routinised behaviour, such as visible swallowing, walking, washing, scratching, basic eating and drinking.
3. Goal-directed artefact manipulation such as moving furniture, combing the hair, applying make-up, tying shoe-laces, conventional eating and drinking.

The first type of practical behaviour, the fortuitous and concomitant behaviours, is evidently culture-independent, though accompanying gestures and interjections may well be highly culture-dependent, depending on sensitivity to pain and on local and personal conventions.

The second type is primarily culture-independent and determined largely by the human anatomy, but secondarily governed by conventions.

The third type is evidently culture-dependent, with highly culture-specific artefacts, ranging from kinds of body decoration in different cultures, including make-up and jewellery in our own, to the instruments of eating: on the one hand, knives, forks and spoons held in certain ways, used in certain combinations and for certain foods; on the other hand the chopsticks and kebab sticks of other cultures, with different manners of gripping. In all these cases there are strong congentions about when these artefacts are to be used (e.g. for food eaten with cutlery) and when not (e.g. finger food).

An interesting feature of these non-semiotic but convention-constrained behaviours is that they can actually also be used semiotically, namely when one of the conventions is deliberately or perhaps involuntarily infringed, and thereby 'makes a statement' about or is interpreted as conveyint a personal attitude of maintaining social distance from a certain group, or even as insulting members of that group.

Further, the artefacts which are manipulated in certain contexts – for example, tying shoe-laces – may have semiotic import in other contexts, e.g. in choosing to wear laced shoes as opposed to sandals.

### 3.1.2 Semiotic behaviour

There are many kinds of gesture-related semiotically functional behaviour, which a comprehensive ontology of gesture needs to relate to each other and to gesture:

1. *Posture*, a holistic configuration of the body, such as the so-called 'Gothic s-curve' of Hogarth's caricatures (cf. Figure 1), or putting body weight on one leg and crossing or moving the other (German has a convenient term for this: the *Standbein-Spielbein* posture).
2. *Orientation*, positioning the entire the body to be facing, near-facing or with the back to the interlocutor.
3. *Distance*, positioning the body to be far from, close to, or very close to the interlocutor (as with Vladimir Nabokov's tragi-comical Timofey Pnin's culturally inappropriate Russian behaviour in the USA).
4. *Clothing* is communication: like dialect, sociolect, style and register in speech and text, clothing has a conspicuous semiotic function indicates, on the one hand, membership of regional origin, social status, formality and activity, and individuality on the other.

The most elaborated approach to categories of this kind is Hall's Proxemic Theory (Hall 1959). The types of semiotic behaviour listed here convey pragmatic information, for example:

1. *Posture* can be interpreted as deferential, threatening, or sexually suggestive.
2. *Orientation* can be interpreted as attentive or as impolite.
3. *Distance* can be interpreted as rejection or as intimacy.
4. *Clothing* can be interpreted as acceptance or rejection of group norms.

Like other forms of verbal and non-verbal communication which convey pragmatic information they can lead to positive and negative attitudinal reactions, and they can be heavily sanctioned if used too non-conformistically: by ostracism, by legal action, or in some situations by reactions of flight or even violence.

*Figure 1: William Hogarth: Mariage à la mode, 1: Le contrat de mariage. National Gallery, London.*

## 3.2 The sociolinguistics of gesture

The study of gesture has always formed a part of discourse analysis, whether in the context of the traditional rhetoric of gesture in public speaking, or in conversation analysis. There have been isolated observations that, for instance, the manual gesture consisting of a loop formed with thumb and index finger has different meanings in different parts of Europe, from "zero" to "very good" to an insultingly obscene metaphorical gesture. There have also been numerous studies of sign languages, mainly the signing of the hearing-impaired and, to some extent of professional gestural communication, such as that of a sports umpire or referee, and in various acoustically unfavourable environments such as the floor of the London stock exchange, or the situation of the bat-man who directs an aircraft into parking position, or semaphore signalling in the navy in earlier days. Much of this kind of gestural communication can be generalised as *teleglossic* communication, to which waving on the one hand and telecommunication registers on the other (Gibbon 1985; Gibbon & Kul 1986) also belong.

But there has been little overall systematisation of the sociolinguistics of gesture in terms of the classic dimensions of language variation. This kind of classification needs to be addressed in a comprehensive ontology, for example with the following dimensions:

1. *Idiolectal*: personal gestural habits, including quirks and twitches.
2. *Regional*: cultural and dialectal variation in gesture, both local and global, such as inter-communicator contact with hand-shaking in Western Europe vs. avoidance of inter-communicator contact with hand-on-heart in parts of West Asia and palm-to-palm 'prayer-like' greeting in India.
3. *Social*: sociolectal variation indicating socio-economic group identity, for instance in greeting gestures, behaviour at table.
4. *Functional*: gesture register based on activity and occupation, often in acoustically unfavourable environments, such as the sports, stock exchange and airports.

The topic of language variation immediately suggests a further issue: gestural universals. It is well known that there are differences along the dimensions mentioned above, but it is not so clear what might be universal and what might be specific. A thorough discussion is not possible in this context, but the following basic tenets are proposed:

1. *Universals*:
    1. All communicator communities use communicative gestures.
    2. Dialogue act functions of gestures are universal, but not forms of gestures.
2. *Specifics*:

1. Vocabulary-related gestures (emblems) culture-specific and environment-specific.
2. The forms associated with universal functions are culture-specific (cf. acquiescence and disagreement by different head movements in different cultures).

# 4 Toward a gesture ontology

## 4.1 Ontologies as taxonomies and mereologies

The term 'ontology' is originally the study of what is, the science of being. In the past twenty years or so it has been specialised in the fields of Artificial Intelligence and Information Retrieval to have a meaning which corresponds roughly to what we know as lexical semantic relations in linguistics (cf. Cruse 1987), and used for classifying concepts in both general and specific technical domains.

Without going into the vast literature on ontologies for library classification, expert systems, the semantic web and many other purposes, it may be noted that the semantic relations involved in current ontologies are relatively simple in comparison with the lexical semantic relations used in lexicology and lexicography. A basic partition of lexical semantic relations will be adhered to in the present discussion: between *paradigmatic* (classificatory, taxonomic, is-a, has-prop) relations and *syntagmatic* (compositional, mereonomic, part-of, part-part) relations, a clear distinction which is not clearly made in current discussion of ontologies but which is foundational (though often also fuzzily and metaphorically defined) in linguistics.

Ontologies have been developed for many disciplines, including linguistics itself, for example the GOLD (General Ontology for Linguistic Description) model (Farrar & Langendoen 2003). The GOLD model concentrates heavily on morphology and syntax; GOLD does not refer in any detail to the modalities of speech and not at all to multimodal communication or to the signing languages of the hearing-impaired. An initial approach to formulating conditions on an ontology for prosody was formulated by Gibbon (2009).

The more general field of multimodal communication remains almost virginal in this respect (but cf. Gibbon 2005). The following sections are intended as a move towards initiating gesture studies into the ontology culture.

## 4.2 Paradigmatic relations: the taxonomy of gesture

Paradigmatic relations are relations of similarity and difference defined in terms of sets or properties, and are generally represented in linguistics as classificatory trees (as in Feature Geometry in phonology, or inheritance lexicons in computational linguistics) or as feature matrices or feature bundles in phonology, or attribute-value matrices in formal syntax and semantics. In mainstream theories, paradigmatic relations in semiotic domains relate the *signifiant* (the material form of the sign) to the *signifié* (the referent of the sign). In the terminology used here, paradigmatic relations hold for modalities and for meanings, e.g. for multimodal gesture complexes, for their functions in communication, and for combinations of these.

A full account of gesture and speech, involving all possible modalities, will be highly complex and far beyond the scope of the present contribution. A selection of visual modality complexes (sets of pairs of motor articulators and the visual sense organ) are shown in Table 1 as an (incomplete) matrix. Gesture articulators are represented as a hierarchy of attributes, sample values of which are contained in the cells of the matrix. The hierarchy is incomplete, in the sense that a much greater granularity is required for a full specification, with explicit mention of individual fingers, right and left hand, elbow, knee, regionsof the head and face, full IPA specifications of speech articulators, and so on.

Gestures do not only generate a visual output. Most of the gestures of speech, which generate an acoustic output, are hidden from view; the exceptions are lip and (to a lesser extent) jaw movements. In gesturing, any or all of the senses may be involved:

1. *Visual*: the gestures categorised in Table 1.

2. *Acoustic*: speech; finger snipping, clapping, stamping, as well as speech surrogates such as whistling, drumming. Music also falls into this general category.
3. *Haptic* (*tactile*): hand-shaking, back-slapping; hugging, kissing, caressing; erotic contact.
4. *Olfactory*: voluntary or involuntary smell (scent; gift of flowers; food and drink, body odour as a fortuitous consequence of proximity).
5. *Gustatory*: voluntary or involuntary taste (food and drink; consequence of erotic contact).

*Table 1: Partial matrix for characterising a taxonomy of functional gesture types mapped to coarse-grained specifications of gesture articulators (for visual inputs only).*

| **Articulators (body areas)** | | | | **Functional multimodal complexes** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | **Phatic** | | | **Speech act** | | | **Appraisive** | |
| | | | | **Wave** | **Salute** | **Air kiss** | **Agreement** | **Don't know** | **Obligation** | **F-sign** | **Thumbs-up** |
| *Head* | *Skull* | | | | | | nod | | | | |
| | *Face* | *Eyebrows* | | | | | | raise | | | |
| | | *Eyes* | | | | | | | | | |
| | | *Nose* | | | | | | | | | |
| | | *Mouth* | *Lips* | | | purse | | | | | |
| | | | *Tongue* | | | | | | | | |
| | | | *Velum* | | | | | | | | |
| | | | *Larynx* | | | | | | | | |
| *Upper* | *Shoulders* | | | | | | | shrug | | | |
| | *Arms* | | | raise | raise & bend | | | | | | |
| | *Hands* | | | flap | | | | | clasp | palm inward | |
| | *Fingers* | | | | extend | | | | | extend & spread index & mid | |
| *Lower* | *Hips* | | | | | | | | | | |
| | *Legs* | | | | | | | | | | |
| | *Feet* | | | | | | | | | | |
| | *Toes* | | | | | | | | | | |
| *Orientation* | | | | facing | facing | facing | | | | | |
| *Posture* | | | | | | | | | | | |
| *Distance* | | | | | | | | | | | |
| *Participants* | | | | >1 | >1 | 1 | 1 | 1 | 2 | >1 | >1 |

## 4.3 Syntagmatic relations: the mereonomy of gesture

The starting point for systematising the parts and combinations of gesture forms is the insight that a single gesture has an identifiable sequence of phases (Kendon 1996), as already noted:

> A gesture is a clearly demarcated symmetrical movement from a rest position via a peak (centre or stroke) back to a rest position.

Gestures defined in this way are primarily *atomic gestures*, which are segments in the *stream of gestures* in the same sense that morphs in the *stream of speech* are segments. The structure of an atomic gesture is analogous to that of the sonority curve of a prototypical CVC syllable in speech: from a well-demarcated low sonority initial consonant through a high sonority vocalic segment to a

low sonority final consonant. Siince gestures, unlike syllables per se, have meanings, they are more analogous to monosyllabic morphs (realisations of morphemes) than to syllables, and a fortiori, at an appropriate level of abstraction, to inventarisable morphemes. Atomic gestures may thus be said to form a basic vocabulary which is formally comparable to the vocabulary of morphemes of spoken language.

Then the issue of syntax arises, a number of two compositional issues, centring on two questions: Is there a of gesture sequences? Is there a grammar of gesture synchronisation? The second issue is formally the same as the that of assigning prosody to locutions as a parallel channel with semi-independent forms and functions.

Based on the present strategy of using linguistic analogies, a number of gesture compositionality issues require treatment:

1. *Composition of atomic gestures*: combinatorics of the simultaneously occurring features which represent the paradigmatic relations between atomic gestures. Some examples of these combinations were given in characterising paradigmatic relations between gestures.
2. *Sequential combinatorics* of atomic gestures, including whether gesture sequences are only 'flat' or linear or whether there are perhaps also hierarchically structured sequences; on very general kind of sequence moves from a wave to start a conversation, via the conversational gestures of the actual interlocution, to a wave at the end. That the sign languages of the hearing-impaired have intricate syntax, like the syntax of speech, is well-known.
3. *Synchronous combinatorics* of atomic gestures, both with each other (e.g. waving and smiling at the same time) and with speech (e.g. pointing gestures together with deictic expressions, emphatic gestures together with emphasised words, overall cohesive gesture gestalts coextensive with utterances).
4. *Word vs. sentence combinatorics*, i.e. whether there is a principled distinction between gestural 'sentences' and gestural 'compound words'. For instance, some gestures consist of two distinct movements in sequence, like the Roman Catholic gesture of crossing oneself, with an assimilation effect of not returning to a rest position between the vertical and the horizontal gesture parts. Gestures of this kind could be described as 'gestural words' or alternatively as 'bi-atomic gestures'. This gesture type contrasts with a combination of two independent deictic gestures, e.g. the index finger pointing to a person, then (or simultaneously, using two hands) the thumb pointing to the door, meaning "You, get out!"

Two complementary recent formal studies of gesture syntax are available:

1. CoGesT: A study of the basic combinatorics of type 1 is given in Gibbon & al. (2003), where a formal grammar for hand gestures is provided. A basic distinction is made into Simplex Gestures and Compound Gestures. Simplex gestures are of two types: *2-place static* (where a gesture with a hand configuration is held) and *9-place dynamic* (with specification of Source (Location and Handshape), Trajectory (Lateral, Sagittal and Vertical Direction; Shape, Form, Size and Speed), and Target (Location and Handshape). These attributes are represented as a vector, which may be enhanced with specifications for two-member gestures (e.g. symmetric, where hands make mirror image movements, or parallel, where hands make the same movement) and indicators for the left or right side of the body for paired members, e.g. left or right hand. The model also has a specification for iterative gestures such as waving.
2. MURML: A set of XML conventions (Wachsmuth & Kopp 2002) for representing gestures in a robotics context, with specifications for Timeline, Symmetry, HandShape, PalmOrientation, ExtendedFingerOrientation, HandLocation, ShoulderLocation, CentreLocation, Start, Direction, Distance. The MURML specification has very many more details for specification of attributes of the hand than can be discussed here.

The basic Source-Stroke-Target structure is represented differently in the two approaches: in CoGesT, the structure is represented directly by the Source-Trajectory-Target triple, while MURML uses a Source-Direction-Distance format which is more convenient for calculations in the robotics environment in which it is located, but does not give the shape of the trajectory.

The third compositionality issue, synchronisation with the utterance, is handled by the timeline in the MURML notation, but requires more detailed temporal models for a full description. A suitable basis for an explication of gesture synchronisation, whether for speech or non-speech gestures, is the notion of *Time Type* (Gibbon 2006), representing levels of abstraction from physically measurable time:

1. *Categorial Time*: time is specified simply an abstract property or category, such as *duration* or concatenation representing sequences in linguistic descriptions.
2. *Relational time* (*rubber time*): time is specified as precedence and overlap relations as used in Autosegmental Phonology, and formalised in van Benthem's Event Logic and in Allen's Interval Calculus (cf. Carson-Berndsen 1998).
3. *Absolute time* (*clock time*): time is specified as a set of measuring points, as in recordings and annotations of digitised audio and video signals.

The variety of temporal relations between gestures and speech is best represented at the level of relations, for instance using Allen's Interval Calculus. The Interval Calculus defines the thirteen possible relations between two intervals *X* and *Y* (see Figure 1). Using this approach, a gesture can be defined as a pair of a *movement* and an *interval*, $G = < M, I >$, and gesture synchronisation can then be defined as a relation between the intervals *X* and *Y* in a set of such gestures: $SYNC(G_X,G_Y)$.

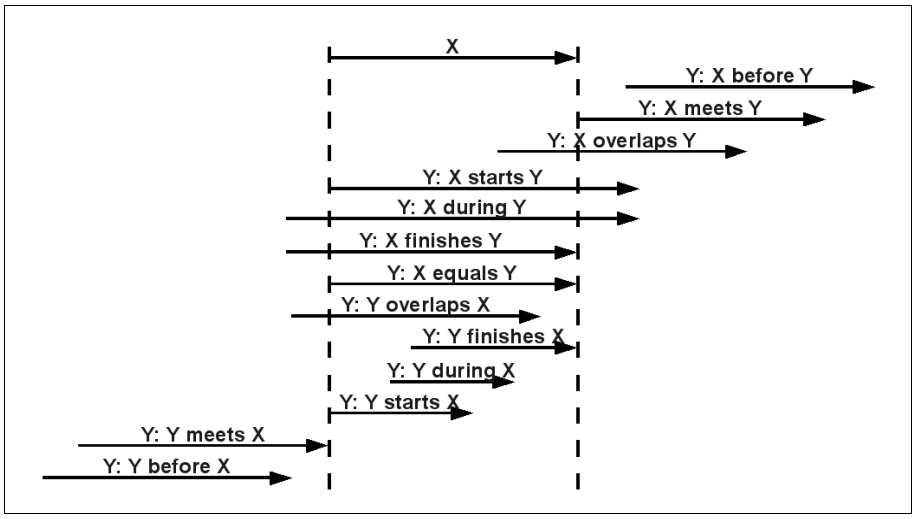


*Figure 2: Illustration of interval relations in Allen's Interval Calculus.*

Carson-Berndsen (1998) has shown how interval and event structures of the kind shown in Figure 1 can be formalised within the Time Type framework as finite state transducers which map between Time Types. Thies (2003) has shown empirically that for certain types of gesture there is a displacement relation: the synchronisation relation between a hand gesture and an associated word constituent is typically, in terms of Allen interval relations, either $OVERLAPS(G_{HAND},G_{WORD})$, $BEFORE(G_{HAND},G_{WORD})$ or $MEETS(G_{HAND},G_{WORD})$.

## 5 Summary, conclusion and outlook

A linguistic approach is taken to the characterisation of gesture, usiing principles for modality identification, segmentation, categorisation and composition. A distinction was made on semiotic grounds between gesture forms and gesture functions, and a specification of both paradigmatic (classificatory, taxonomic) and syntagmatic (compositional, mereonomic) relations between gestures and between gestures and speech was outlined.

It is tentatively concluded that gestures function in the same way as speech, and can be classified in fundamentally the same ways. Conversational gestures are not a system *sui generis*;

they can be usefully modelled by established methods of linguistic description and representation. The relations of the forms of gestures to each other and to speech can be modelled in the same way as the relation between the locutions and prosodies (intonation, stress patterns, duration patterns), or between the tiers in Autosegmental Phonology, namely with an Interval Calculus or an Event Logic. Models for these approaches can be provided with operational interpretations within the Time Type framework in terms of finite automata.

A comprehensive formalisation of gestures and associated forms of communication using the techiques discussed here – in particular an ontology for paradigmatic relations and for syntagmatic relations, and an operational modelling strategy for the ontology – has not so far been attempted. Consequently, we have very much to look forward to in connection with this field in the future.

## Bibliography


Browman, Catherine & Louis Goldstein (1992). Articulatory Phonology: An overview. *Phonetica* 49:15-180.

Carson-Berndsen, Julie (1998). *Time Map Phonology: Finite State Models and Event Logics in Speech Recognition*. New York: Kluwer Academic Publishers.

Clements, G.N. (1985). The geometry of phonological features. In: E. Colin Ewen & John Anderson, eds., *Phonology Yearbook* 2.

Cruse, D. Alan (1987). *Lexical Semantics*. Cambridge: Cambridge University Press.

Farrar, Scott & D. Terence Langendoen (2003). A linguistic ontology for the Semantic Web. *GLOT International*. 7, 97-100.

Gibbon, Dafydd (1996). *Perspectives of Intonation Analysis*. Bern: Lang.

Gibbon, Dafydd (1985). Context and variation in two-way radio discourse. in: Charles A. Ferguson, ed. *Discourse Processes*, 8,4:391-420.

Gibbon, Dafydd (2005). Prerequisites for a Multimodal Semantics of Gesture and Prosody. In: Harry Bunt, ed., P*roceedings of the International Workshop on Computational Semantics* 6.

Gibbon, Dafydd (2006). Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: Sudhoff, Stefan & al. *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter., pp. 281-209.

Gibbon, Dafyddd (2009). Can there be standards for Spontaneous Speech? Towards an Ontology for Speech Resource Exploitation. In: Shu-Chuan Tseng, ed. *Linguistic Patterns in Spontaneous Speech*. Language and Linguistics Monograph Series A25. Taipei, Taiwan: Institute of Linguistics, Academia Sinica.

Gibbon, Dafydd, Ulrike Gut, Benjamin Hell, Benjamin, Karin Looks, Alexandra Thies, Thorsten Trippel (2003). A computational model of arm gestures in conversation. *Proceedings of Eurospeech 2003*:813-816.

Gibbon, D. & Kul, M. (2006). Economy strategies in English and Polish text messages as examples of channel constraints. Presentation at Poznań: Poznań Linguistic Meeting. http://wwwhomes.unibielefeld.de/~gibbon/Docs/gibbon-kul-smschannelconstraints04.pdf

Hall, Edward T. *The Silent Language*. Garden City, N.Y.: Doubleday, 1959.

Kendon, Adam (1996). An agenda for gesture studies. *The Semiotic Review of Books*, 7.3:7-12.

Lakoff, George & Mark Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

McNeill, David (1992). *Hand and Mind*. Chicago: University of Chicago Press.

Pittenger, Robert E., Charles F. Hockett and John J. Danehy (1960). *The First Five Minutes*. New York: Martineau.

Thies, Alexandra (2006). First the Hand, then the Word: On Gestural Displacement in Non-Native English Speech. MA equiv. thesis, Universität Bielefeld.

Wachsmuth, Ipke & Stefan Kopp (2002). Lifelike gesture synthesis and timing cor conversational agents. In: Ipke Wachsmuth & Timo Sowa, eds. *GW 2001, LNAI 2298*:120-133. Berlin: Springer Verlag.