ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/228508573

Why gesture without speech but not talk without gesture?

ARTICLE · JANUARY 2011

READS
28

2 AUTHORS:

Nicla Rossini
University of Pavia
14 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Dafydd Gibbon
Bielefeld University
108 PUBLICATIONS **712** CITATIONS
SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Nicla Rossini
Retrieved on: 04 February 2016

Why gesture without speech but not talk without gesture?

Nicla Rossini [1], Dafydd Gibbon [2]

[1] Istituto Internazionale per gli Alti Studi Scientifici "E. Caianiello", Vietry sul Mare, Italy
[2] Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Germany
nicla.rossini@unipv.it gibbon@uni-bielefeld.de

Abstract

Explanations of the phylogeny and ontogeny of gesture and speech require not only an understanding of empirical gesture types and scenarios, but also of formal properties of the relations between visual and vocal gesture in a grammar of gesture. Initially we point out that speech (and most obviously prosody) is acoustically transduced gesture, a phonetic truism, and then develop a movement vs. gesture (MSC-GSC) model of the relations between pre-semiotic and semiotic visual and vocal gesture, and a four component linear-feature-timing-realtime (LFTR) grammar model, and apply well-tried phonetic measures to elementary beat gestures and their function in establishing rhythmic coherence, as a first step in formal evolutionary reconstruction.

Index Terms: gesture, prosody, speech, language evolution.

1. Introduction

In this contribution we presuppose an intuitive understanding of the terms 'gesture' and 'speech', and discuss prerequisites for explaining the phylogenetic and ontogenetic origins of visual and vocal gesture. Visual gesture is not necessarily *sui generis*. Speech is also gesture: speech sounds are acoustic transductions of gestures of the vocal tract, a fundamental phonetic fact documented in the International Phonetic Alphabet; cf. also Bolinger's view [1] that language is embedded in gesture. Physically, speech gestures are no different from (though vastly more complex than) other acoustic gesture transductions: lip-smacking, foot-stamping, finger-snipping, hand-clapping.

Not only is detailed empirical evidence required for plausible explanations, but also clear formal foundations. Without integrative, formal and quantitative models discussions remain in the realm of insightful but anecdotal or metaphorical speculation. McNeill's 'growth point' metaphor [2], for example, is a highly productive contribution to gesture theory, but difficult to capture precisely. Formal models are of course not ends in themselves, but instruments for avoiding misunderstanding.

We introduce three contributions to the language evolution topic and claim that each is relevant for both visual and acoustic gesture (with a critique of previous approaches in Section 3):

1. an integrative visual-acoustic gesture model, distinguishing semiotic prosody-like 'gesture' and articulated 'speech' configurations (GSC), as opposed to non-semiotic body 'movement' and 'sound' configurations (MSC) in general interaction: the MSC-GSC model (Section 2);
2. prerequisites for evolutionary explanation in a four-level model of gesture and speech form which capture gesture sequencing, internal structure, temporal relations and real-time events: the LFTR model (Section 4);
3. an empirical approach to real-time beat gesture modelling with measures borrowed from phonetics, for three deliberately selected, very different gesture scenarios, with similarities which will require explanation in any discussion of phylogeny and ontogeny (Section 5).

Finally the significance of these results for the 'ex pede Herculem' reconstruction problem of extrapolating synchronic analysis to language ontogeny and phylogeny is discussed, in order to complement our illustrative model with a predictive theory.

2. The MSC-GSC model

2.1. Basic distinctions

We make four basic distinctions:

1. *modality* (motor-sensory output-input channels of human communication) and *medium* (intervening visual, acoustic, tactile, olfactory, gustatory and technical channels); cf. Gibbon et al. [2000];
2. physical *modality dimensions* and semiotic *functional dimensions* of a gesture or speech sign;
3. *pre-semiotic* or *non-semiotic* body movement and sound configurations (MSC), and *semiotic* gesture and speech configurations (GSC) on an evolutionary scale from MSC to GSC interaction;
4. *potentially iterative* (linearly recursive) pre-speech semiotic gestural, paralinguistic and prosodic vocal events, and *potentially centre-recursive* (with hierarchical embedding) visual and vocal gesture events.

Table 1: The Movement-Sound-Configuration (MSC) to Gesture-Speech-Configuration (GSC) evolutionary scale.

	MSC	**GSC (semiotic)**	
structure	*arbitrary*	*iterative*	*centre-recursive*
visual modality	actions	gesture	signing (sign languages)
acoustic modality	vocal sounds	vocalisations,paralinguistic& prosodic features	speech

EVOLUTION

This model of gesture-sound relations, the *MSC-GSC Model*, is visualised in Table 1. The evolutionary relation represented by the model does not mean that MSC stages of evolution are *supplanted* by later GSC stages, but that MSC stages are *supplemented* by later GSC stages in an evolutionary maturation process of increasing complexity.

The MSC-GSC model suggests that gesture and prosodic vocal gestures precede but are not directly comparable with articulated locutionary speech: the appropriate comparandum for speech is signing; both are more complex than prosody and conversational gesture. The appropriate comparandum for gesture is prosody: fundamentally fairly simple linear or flat-hierarchy sequences (e.g. iterations of *preparation-stroke-retraction* and variants of this), as in intonation, were shown by Kendon [4]. Prosody and gesture are not centre-recursive

but iteratively recursive, except perhaps when driven by locutionary configurations; cf. Fricke [5]. Gestures and prosody share indexical (deictic, idiolectal) and iconic (temporally ordered, onomatopoeic and phonaesthetic) elements; cf. Gibbon [6]. Some aspects of gesture (emblems, icons) overlap with small subsets of vocabulary in articulated locutionary speech.

If speech and gesture are both inserted systematically into a multi-dimensional Communication Space defined by their ranges of form, structure and function, it becomes clear that prosody, paralinguistic gestures and vocalisations share many indexical and phatic functions with manual and brachial gestures, postural movements, movements of the pelvis and legs. Some gestures are simultaneously transduced by impact or vibration into sounds: lipsmacking, air-kisses, finger-snapping, hand-clapping, thigh or table beating, foot-stamping. Other gestures are transduced into tactile contact: hand-shaking, stroking, embracing. Some of these transduced gestures are simultaneously visual, acoustic and tactile: back-slapping, 'high five'. Some tactile gestures are accompanied by gustatory, olfactory and erotic sensations: hugging (olfactory), kissing (gustatory, olfactory), both perhaps also erotic. The modality dimensions also include technically transduced gestures in different media, such as typing and mouse-moving, morse code transmission, music and painting. Vocal and instrumental music production are also acoustically transduced semiotic gesture, but, like signing and speech, more complex than conversational gesture, being in principle centre-recursively rather than merely iteratively structured.

Consequently, we suggest that visual and vocal gestural communication evolved simultaneously and interdependently, with simple gestures such as beats (batons) and the function of communicating rhythmic coherence as the starting point for gestures in deliberate communication (cf. Section 5). The MSC-GSC model also suggests that the phylogenetic and ontogenetic maturation of MSC patterns into functional semiotic GSC patterns parallels the maturation of the central nervous system into its adult state in *homo sapiens sapiens*.

3. Gesture-speech evolution

The views represented in the literature may be summarised as separate hypotheses:

1. The Prioritary Hypothesis: *gesture came before language*, Arbib [7], according to which language started gestural and stayed gestural, retaining the strong iconicity and indexicality of gesture in pronunciation, morphology and grammar, and with incomplete semiotic arbitrariness.
2. The Multimodal Hypothesis: *gesture and speech are both language*, proposed by McNeill [2].
3. The Segregative Hypothesis: *speech and gesture are quite different*, with the special case of *speech is the sole form of language*, as maintained by the Chomsky school (passim) which is unique to *homo sapiens sapiens*.

We note that Chomsky's newer claim that recursion is the only trait that distinguishes human from animal communication; cf. Hauser, Chomsky and Fitch [8] may apply, logically, to all three hypotheses.

The principles outlined in the present study and their application to the phylogeny of commnication did not develop in a vacuum, but are partly distilled from previous studies dating back to antiquity, and pursued with increasing intensity in recent centuries; cf. Kendon [9]. The 18th century enlightenment philosopher Condillac speculated that language evolved as a social convention from a previous system based on visual gestural signals. But speculations on the evolution of language became so numerous that the Linguistic Society of Paris banned the topic as unanswerable in 1886.

After a long period of silence, hypotheses on the evolution of language were revived in biological contexts by Lenneberg [10] and have more recently found a new empirical basis thanks to Magnetic Resonance (MR) experiments, e.g., Arbib [7]. Stimulating contributions to the field also come from studies of the development of child language, cf. Slobin [11], gesture studies, Kendon [9], McNeill [2], sign language studies, Emorrey [12], Kimura [13], and ethological studies focusing on animal communication, especially primate communication, by Tomasello and Camaioni [14] ,and bird song, by Gentner et al. [15].

3.1. The gesture-first hypothesis

The MSC-GSC model appears to suggest that Condillac's 'gesture first' hypothesis is *prima facie* trivially wrong. But the gesture first hypothesis simply turns out to be massively incomplete rather than wrong, ignoring paralinguistic vocal gestures (e.g. voice quality, pragmatic properties of intonation) and prosody (e.g. rhythm, structural intonation).

The gesture first hypothesis is revived in the recent literature, e.g. by Givón [16], supported by the interpretation of ontogenetic patterns in language evolution, with a particular interest in the emergence of language in children as a model for the phylogenetic evolution of language, because gesture appears to emerge before speech in infants. The assumption is further supported by empirical data on the anatomy of previous species of *homo* which show that phonation would be impeded; cf. Lenneberg [10].

Support also comes from studies of sign language in apes; cf. Tomasello and Camaioni [14]. Scholars maintaining the gesture first hypothesis infer, as a consequence of ontogenetic development and the structure of the vocal organs, that the vocal-auditory channel must be a secondary specialisation enabling articulated communication. More evidence comes from fMRI investigation of innately deaf subjects, who have been shown to present an activation of the auditory cortex when signing or perceiving a sign; cf. Emorrey [12].

An interesting indirect argument for a close link between visual gesture and vocal gesture comes from the relatively recent discovery of mirror neurons in primate brains, particularly in Broca's area, which suggests a close link between communicative actions and language, and is supported by the observation that in synchronic utterances synchronised and co-expressive manual gestures usually precede the concurrent speech; cf. McNeill [17]; Thies [18].

However, there are two serious defects in in the 'gesture first' view: incompleteness and restrictedness.

First, the coexistence of gestures and semiotically functional pre-speech paralinguistic and prosodic vocalisations is not taken into account. It is well known that many sounds with semiotic function are uttered by babies before the onset of structured speech, cf. Slobin [11], and by many species of animals. The MSC-GSC model suggests that speech (and signing) have evolved ontogenetically and phylogenetically by gradual maturation processes in which the complexity of semiotically meaningful sounds increases with time. The gesture first view may perhaps be interpreted as an incomplete segment of the MSC-GSC model.

Second, the structuring of conversational gestures is highly restricted in face-to-face communication in comparison with articulated locutionary speech itself, and (like prosodic and paralinguistic features and vocalisations) contributes practically nothing to understanding the gist of a conversation when observed by non-participants; cf. Feyereisen et al. [19].

Overlaps are small: icon, emblem and metaphorical gestures are comparable with simpler elements of signing, with interjections auch as 'oh', 'ouch', 'mm', 'er', for example, and perhaps with emphatic prosody such as lengthening and voice lowering in 'biiiiig' etc. Gesture-speech relationships are more complex than simple evolutionary precedence would allow.

3.2. The multimodal hypothesis

The "multi-modal hypothesis" of McNeill [2] maintains that language could not evolve in any other form than the one we can observe currently, i.e. as a multi-modal system. McNeill's argument is that this is the result of different and concurrent types of thought: analytic and articulated thought gives birth to speech, and global and synthetic thought called 'imagery' gives rise to manual gesture. Because these two types of thought are simultaneous they tend to collide with one another and create what McNeill [2], [17] calls a self-organised system. The opposition of imagery and linguistic thought is the foundation of ideas that are conveyed by both speech and gesture that are parts of a unitary system. According to McNeill, speech and gesture are not redundant with respect to one another, but serve different and complementary functions.

A number of dedicated experiments have been designed in order to assess the plausibility of the different hypotheses. Among the first of these is Rimé's [20] finding that co-verbal gestures are produced even under conditions of blocked visibility (e.g. telephone conversation), when subjects know that their gestures do not reach the interlocutor, indicating that co-verbal gestures may have non-communicative functions. Feyereisen et al. [19] have shown that the content of a conversation cannot be understood with the sole observation of gestures, thus also undermining the concept of autonomous communicative function for gesture.

We regard the binary dialectic imagery/thought approach as too simple and metaphorical. We also take a multimodal approach, but the MSC-GSC model suggests that gesture and speech relations are not directly comparable. Therefore *of course* gesture and speech are, trivially, complementary and serve different and complementary functions. But gestures do not serve different and complementary functions in relation to the category of vocalisations, paralinguistic features and prosody. On the contrary, they share the same functions, except for some modality-specific scenarios. We therefore strongly disagree with the claim that visual and vocal gesture serve different and complementary functions in any straightforward binary sense, though gesture and prosody *may* be instrumentalised to serve different functions in specific utterance instances (as in irony or double-bind situations).

4. Modelling gestures

Descriptions of the 'grammar' of gesture are still rather fuzzy by linguistic standards, despite initial work by Gibbon & al. [21], Rossini [22], Müller [23], Bressem & Ladewig [24]. However, an adequate formal model of gesture grammar is still lacking. We introduce a 4-component grammar model, covering a linear (L) component, a feature (F) component, a timing (T) component and a realtime (R) component.

4.1. Linear grammar (L) component

We start with the linear patterns used in many approaches:

1. kinesic unit, e.g. Birdwhistell [25]: the complex of hand movements occurring between two phases of rest position;
2. gesture phrase: the single gesture as commonly perceived by a naïve public; gesture phase: a "kineme" of the gesture phrase; *onset excursion offset*; cf. Kendon [9],
3. *preparation* (*pre-stroke hold*) *stroke* (*post-stroke hold*) *retraction/reposition*', cf. McNeill [2], [17].

Although the stroke is seen as the essential core gesture phase, recent studies suggest that all phases contribute to gesture functionality, cf. McCullough [26], and that pre-stroke and post-stroke holds also make essential contributions to meaning. The archetypal linear patterning is found in the rhythms of beat gestures and of accents in speech, and we use phonetic methods from prosody studies to examine them (cf. Section 5). Linear patterns need to be supplemented by other structures, however.

4.2. Feature structure (F) component

The second component captures the internal complexity of gestures using familiar linguistic and phonetic techniques. Gibbon et al. [21] decompose linear template elements into simultaneous attributes and values, including hand shape, position in a 3-dimensional space around the body, and trajectory shape, size and speed. In 3-dimensional space the trajectory shape may be monotonic, non-monotonic, or iterative, or contain segments with any of these properties. Bressem and Ladewig [24] decompose hand properties into feature structures.

4.3. Timing (T) component

The third component is timing in the visual stream, the acoustic stream, and in the synchronisation of the two. The simple attribute-value model, complex as it is, fails with timing: gesture components simply do not occur in tightly co-ordinated temporal chunks. An appropriate model is that of autosegmental phonology, in which simultaneous and partly independent streams, 'tiers', of articulatory gestures (e.g. lip-rounding, fronting and raising of the tongue, voicing, tone) are modelled as sequential and overlapping events. The universal relations involved were formally modelled by Allen [27], shown in Figure 1.

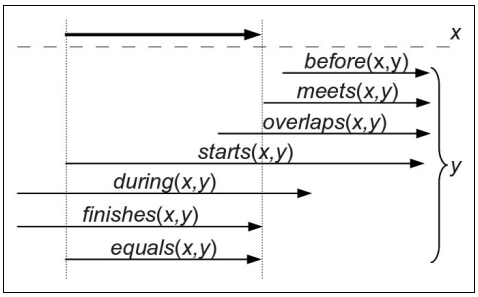

Figure 1: Temporal relations in the Allen interval calculus: 14 relations in all, i.e. the 7 relations shown and their inverses.

The Allen relations were originally developed to model temporal relations exhaustively in any domain and to serve as the basis for inference algorithms in various disciplines from computational linguistics through robotics to railway and airline safety systems. Consequently, they also accurately model sequential and overlapping visual and vocal gestures of all kinds. It has been shown by Kay [28] that an appropriate processing model for temporal relations is a generalised finite-state transducer with as many tapes as there are tiers to model. Berndsen's Time Map Phonology calculus [29] expands Kay's approach to a cascade of feature-annotated transducers.

The Allen relations apply to complex gestural grammar in the following way: gesture components on the same tier (i.e. values of the same attribute) can only occur in the relations

before(x,y) and *meets(x,y)*. Co-expressive gesture components on different tiers (i.e. values of different attributes) can enter into any of the relations. As already noted, there is a constraint that the stroke of a gesture occurs before or concurrently with a gesture, but not after. This constraint is also found by Rossini [30], [31] in congenitally profoundly deaf subjects educated with oral Italian: even in these cases, where subjects have never been exposed to visual-acoustic synchronization patterns, the stroke of the gesture tends to either precede or synchronize with the most prominent syllable in speech. The constraint is easy to express formally with Allen relations:

$$\neg\ (before(word,stroke) \vee meet(word,stroke))$$

The beat gestures which we address in the following section involve the *before* relation with each other, and, since they are rather fast, mainly the *before* and *during* relations with prosodic accentuation.

4.4. Real-time interpretation (R) component

The fourth component goes beyond a purely relational 'rubber timing' with indeterminate interval length, and calls for interpretation of the relations in real-time terms as 'clock timing', such as the actual lengths of a beat-baton interval and of an accented syllable. The relation of a beat-baton gesture (cf. Section 5) to accentuation in speech may be of the order of 200-700 ms, whereas the same relation for iconic or emblem gesture intervals and words may be of the order of >1 seconds, cf. Thies [18]. Relations alone are not maximally informative.

5. Gesture-prosody scenarios: beats

We select beats for empirical examination of a key gesture type with a very basic semiotic function of indicating rhythmic coherence, like accentuation in speech, which we claim occurs early in the evolution of semiotic functions of gesture and speech, and demonstrates a transitional behaviour between MSC and GSC patterns, with simple '*before(x,y)*' temporal relations. Beats may therefore serve as a starting point for further empirical investigation of gesture and speech synchronisation; the extensive work on visual-vocal synchronisation, e.g. by Kendon [4], Thies [18], McNeill [17], cannot be reviewed here, but cf. Sections 5.2 and 5.3.

5.1. Ega story-telling

The first scenario to be briefly outlined is story-telling in Ega, (Niger-Congo, Kwa, Côte d'Ivoire, ISO 639-3: *ega*). Figure 2 is a still image from a video recording of a narration, showing the peak excursion of of a medium movement right-hand beat gesture. In this study, beats are analysed as autonomous patterns, not in relation to accents in speech: synchronisation with accents is a separate topic.

The narration is puncuated by numerous iterated beat gestures (in a 'before' relation) at a rather fast rate (a 'heartbeat' rate averaging 85 per minute). Beats have, intuitively, the function of manifesting rhythmic coherence, like prosodic accentuation and rhythm, with 'iterative GSC' functionality. Parallels in this field have been sporadically investigated from Lashley [32] to Cummins & Port [33], for example. A frequently used phonetic measure for regular temporal patterning was selected, the *normalised Pairwise Variability Index*, *nPVI* (cf. Gibbon [6] for details and critique). The *nPVI* is a function of the average absolute differences between durations of adjacent units in utterances, and varies from 30 to 70 for different languages.

The recording was annotated for beat gestures of the hands (with the *Anvil* video annotator). For present purposes, left-hand, right-hand and synchronised beats were combined. Time-stamps were extracted, and the *nPVI* applied. The result, *nPVI*=5, clearly implies extreme regularity, confirming the intuition of regularity, particularly since pauses of various lengths were not excluded. The ranges and standard deviation are rather high, however, which rather belies the *nPVI* value:

min=80ms	*max*=3200ms	*range*=3120ms
mean=770ms	sd=730ms	*nPVI*=5

Figure 2: Still frame with medium amplitude right-hand baton gesture (Ega, chef-conteur Grogba Marc).

A visualisation of the timeline, in which beat annotations on a scale from 1 (small beat excursion) to 4 (very large beat excursion), indicates more complex patterning and different strata of rhythm patterns (Figure 3). Slopes of lines between peaks of given heights are a function of speed of repetition, and flatter slopes after higher peaks tend to indicate pauses and the end of major discourse units. A full analysis of the gesture rhythm in relation to speech rhythm is beyond the scope of the present discussion.

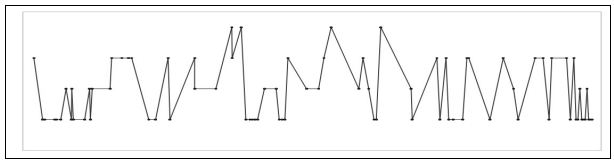

Figure 3: Beat gesture timeline in first 60s of Ega story. Beats are categorised from small (level 1) to very large excursions (level 4); lines connect the beat peaks. Flatter slopes of higher jumps tend to indicate the end of discourse units.

5.2. Multi-tasking: beat actions and speech

In addition to investigating gestural temporal patterns alone, the coordination of speech prosody and gesture has been investigated in multi-tasking experiments. If manual action and speech are completely detached from one another, then any type of experiment involving the execution of both tasks concurrently should be easy to perform. If, instead, either action or speech is dominant in the linkage, one task will necessarily influence the other (but not the contrary).

In a multi-tasking experiment by Rossini [34], [31], participants were asked to read two Italian texts (the first one in prose, the second one in poetry) while repeating a manual rhythmic beat given by an experimenter (cf. Figure 4). For one speaker, the same values were calculated as for the narration scenario:

min=290ms	*max*=359ms	*range*=71ms
mean=320ms	*sd*=18ms	*nPVI* = 5

Comparing the ratios of the duration ranges (3120:71) and the standard deviation ratios (730:18) it is clear that there is much greater beat regularity in the formal experimental

setting. In fact, the mean ratios (770:320), in which the experiment beats are twice the speed of the narration beats, may indicate that a different rhythmic clock mechanism is in place in the experimental scenario. The *nPVI*, on the other hand, with a value of 5 in each case, did not yield the expected difference between scenarios.

Figure 4: Still frame with synchronous reading aloud and rhythm-beating with forefinger.

Table 2: Multi-tasking (M-T) experiment (Rossini [34], p. 445). Column 1: Ctimings of subjects during prose reading. olumn 2: timings during multi-tasking with prose.

	Prose only	Prose M-T	Poem only	Poem M-T
S1	2:30’	2:11’	0:16’	0:20’
S2	3:03’	2:48’	0:28’	0:23’
S3	2:28’	2:10’	0:17’	0:21’
S4	1:30’ (p)	1:23’ (p)	0:26’	0:25’
S6	0:51’ (p)	0:39’ (p)	0:23’	0:23’
S7	1:22’ (p)	1:13’ (p)	0:22’	0:21’
S8	1:25’ (p)	1:15’ (p)	0:20’	0:23’
S9	1:36’ (p)	1:27’ (p)	0:26’	0:21’

The main goal of the experiment was to investigate synchronisation, however. The results of the experiment (see Table 2) show that not only does speech influence the beat in manual action, but that also manual action influences speech in causing disruptions, at least in prose, where the prose only differs consistently from the Prose M-T condition; for further details cf. Rossini [31].

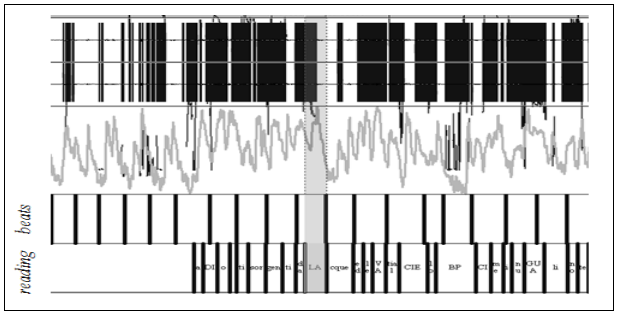

Figure 5: Multi-tasking experiment (Rossini [34], p. 446. Detail of S1’s performance during the first nine seconds of prose reading in multi-tasking.

In particular, it has been observed that a mismatching rhythm in the hand beat influences reading both in terms of speech rate, and in terms of speech disfluency. Figure 5 shows an annotation of a recording with an example of syncopation in both hand beat and speech so as to align the manual downbeat with the accented syllable in speech.

5.3. Speech and gesture in hearing and deaf subjects

Empirical clarification of the autonomy of gestures with regard to speech is an essential prerequisite for explanation of the gesture-speech priorities suggested by the MSC-GSC model (cf. also Section 3).

A good number of studies have confirmed the trend of gestures preceding words in both hearing and deaf-born subjects; cf. Rossini [30], [31]. In particular, data from spontaneous speech in congenitally profoundly deaf subjects educated with oral Italian has shown that even in cases where the subjects have never been exposed to synchronization patterns because of profound deafness, the stroke of the gesture tends to either precede or synchronize with the most prominent syllable in speech. In this case, the speech and gestural production of two profound deaf subjects aged 45 and 47 years, respectively, with no acoustic aid and an acoustic loss of 70db per ear was analyzed. The data thus obtained were described in terms of several parameters such as speech rate, gesture rate, gesture size, point of articulation and locus. We focus on the relationship between the stroke of the gesture and the concurrent speech: all strokes were performed either *in correspondence with* or slightly *before* the accented syllable of the speech flow the gesture occurred with, while no strokes were performed *after* the accented syllable (Table 3); cf. also Section 4.3.

Table 3: Values and percentages of strokes performed by two profound deaf subjects and a hearing interviewer (Rossini, [30], [31])

	Subject 1		Interviewer 1		Subject 2	
Strokes	**values**	**%**	**values**	**%**	**values**	**%**
in corresp.	134	45	42	64	51	42
NS	85	28	4	6	46	38
before	67	22	14	22	23	19
Nd	10	3	0	0	1	1
NS listener	3	1	0	0	0	0
NS pause	3	1	0	0	0	0
NS hes	0	0	5	8	0	0

In particular, the first subject produced 134 strokes out of 302 (45%) in correspondence with the accented syllable, 67 strokes (22%) before the correspodent accented syllable, and 91 (30%) with no speech. The synchronization of 10 (3%) gestures with the corresponding speech was not determinable due to overlapping turns and the conspicuously poor oral proficiency of the participant. The second subject produced 51 strokes out of 65 (42%) in correspondence with the accented syllable, 23 strokes (19%) before the corresponding accented syllable, and 47 (39%) with no speech.

6. Conclusion: gesture-speech integration

We have provided systematic models for the description of functional gesture-speech relations (the MSC-GSC model) and a four-component gesture-speech grammar (the LFTR model). Like most other approaches, we have concentrated on descriptive dimensions of gesture production, ignoring other procedural dimensions such as perception and learning. On the empirical side we have deliberately started with simple beat gestures, which we speculate are located very early on the semiotic evolutionary scale, and we have shown that visual gesture behaviour can be described quantitatively with established phonetic measures taken from prosody studies.

Our empirical investigations show that beat gestures have consistent temporal rhythmic properties across very different scenarios. Future work will need to address more sophisticated rhythm models. The integrated MSC-GSC model permits the formulation of useful hypotheses about pre-speech gesture and

prosody, versus gesture and prosody together with articulated locutionary speech on the other. The four-component LFTR grammar model captures the formal properties of visual and vocal gesture systems in detail.

The motivation for the present contribution was to try to find out what is needed for a coherent evolutionary model of gesture and speech. On the basis of this work we suggest specifically (1) that the communicative structures shown in both the MSC-GSC the LFTR grammar models evolved gradually, (2) that this evolution occurred simultaneously for visual and acoustic gestures, and (3) that rhythmic iterative beat recursion and synchronisation arose at an early state in the transition from MSC to GSC scenarios, as a prerequisite for complex hierarchical recursivity in signing and speech.

In a nutshell: conversational visual gesture and prosodic and paralinguistic vocal gestures developed together; articulated speech and signing developed later. Consequently, gesture and speech, in the usual senses which figure in the literature, are not directly comparable.

7. References

[1] Bolinger, D.L., "Intonation and Gesture", American Speech, 58:156-174, 1983.

[2] McNeill, D., Gesture and Thought, Chicago, Illinois, USA: University Of Chicago Press, 2005.

[3] Gibbon, D., Mertins, I., and Moore, R. [Eds], Handbook of Multimodal and Spoken Dialogue Systems, Dordrecht etc.: Kluwer Academic Publishers, 2000.

[4] Kendon, A., "Some Relationships between Body Motion and Speech. An Analysis of an Example", in Wolfe, A. and Pope, B. [Eds], Studies in Dyadic Communication. Pergamon Press, New York, 1972.

[5] Fricke, E., Grundlagen einer multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen. Europa-Universität Viadrina, Frankfurt (Oder), 2008. (Ms. unavailable, citation from abstract.)

[6] Gibbon, D., "Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data", in Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P. [Eds], Methods in Empirical Prosody Research, 281-209, Berlin: W. de Gruyter, 2006.

[7] Arbib, M. A., Action to Language via the Mirror Neuron System, Cambridge: Cambridge University Press, 2006.

[8] Hauser, M. D., Chomsky, N., and Fitch, W. T., "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?", Science, 298:1569-1579, 2002.

[9] Kendon, A., Gesture: Visible Action as Utterance, Cambridge, Cambridge University Press, 2004.

[10] Lenneberg, E. H. Biological Foundations of Language, N. Y.: John Wiley & Sons, Inc., 1967.

[11] Slobin, D. I., "From ontogenesis to phylogenesis: What can child language tell us about language evolution?", in Langer, J., Parker, S. T. and Milbrath, C. [Eds.], Biology and Knowledge revisited: From neurogenesis to psychogenesis, 255-285, Mahwah, NJ: Lawrence Erlbaum Associates, 2004.

[12] Emmorey, K., Language and Space, in Penz, F., Radick, G. and Howell, R. [Eds], Space, 22-45, Cambridge University Press, 2004.

[13] Kimura, D., Neuromotor mechanisms in human communication, N. Y.: Oxford University Press, 1993.

[14] Tomasello, M., Camaioni, L., "A comparison of the gestural communication of apes and human infants", Human Development, 40:7-24, 1997.

[15] Gentner, T. Q., Fenn K. M., Margoliash D. and Nubaum H. C., "Recursive syntactic pattern learning by songbirds", Nature, 440:1204-1207, 2006.

[16] Givon, T., Bio-Linguistics: the Santa Barbara Lectures, Amsterdam and Philadelphia: Benjamins, 2002.

[17] McNeill, D., Hand and Mind: What Gestures Reveal about Thought, University of Chicago Press: Chicago and London, 1992.

[18] Thies, A., First the Hand, then the Word: On Gestural Displacement in Non-Native English Speech, Bielefeld: SII thesis, Universität Bielefeld, 2003.

[19] Feyereisen, P., Wiele, M. v. d., Dubois, F., "The meaning of gestures: What can be understood without speech?", Cahiers de Psychologie Cognitive / European Bulletin of Cognitive Psychology, 8:3-25, 1988.

[20] Rimé, B., "The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal behaviour", European Journal of Social Psychology, 12:113-29, 1982.

[21] Gibbon, D., Gut, U., Hell, B., Looks, K., Thies, A., Trippel, T., "A computational model of arm gestures in conversation", in Proc. Eurospeech 2003, Geneva, 2003.

[22] Rossini, N., "The Analysis of Gesture: Establishing a Set of Parameters", in Camurri, A.-Volpe, G. [Eds], Gesture-Based Communication in Human-Computer Interaction. 5th International Gesture Workshop, GW 2003, Genova, Italy, April 2003. Selected Revised Papers, 124-131, Springer: Berlin etc., 2004a.

[23] Müller, C. and A. Cienki, "Words, gestures, and beyond. Forms of multimodal metaphor in the use of spoken language", in Forceville, C., Urios-Aparisi, E. [Eds.], Multimodal Metaphor, 297-328, Mouton de Gruyter: Berlin, N. Y., 2009.

[24] Bressem, J. and Ladewig, S. H., "Rethinking gesture phases: Articulatory features of gestural movement?", Semiotica, 184-1/4:53-91, 2011.

[25] Birdwhistell, R. L., Introduction to Kinesics, Washington DC: U. S. Dept. of State Foreign Service Institute, 1952.

[26] McCullough, K-E., Using Gestures during Speech: Self-Generating Indexical Fields, Ph.D. thesis, The University of Chicago, Chicago, Illinois, 2005.

[27] Allen, J. F., "Maintaining knowledge about temporal intervals", Communications of the ACM, 26 November 1983, 832-843, ACM Press, 1983.

[28] Kay, M., "Nonconcatenative Finite-State Morphology", Proc. Third European ACL Conference, 1987.

[29] Carson-Berndsen, J., Time Map Phonology: Finite State Models and Event Logics in Speech Recognition, Dordrecht: Kluwer Academic Publishers, 1998.

[30] Rossini, N., Gesture and its cognitive origin: Why do we gesture? Experiments on hearing and deaf people, Ph.D. thesis, Università degli Studi di Pavia, 2004b.

[31] Rossini, N., Language in Action. Reinterpreting Gesture as Language, IOS Press: Amsterdam, etc., in press.

[32] Lashley, K. S., "The problem of serial order in behavior", in Jefress, L. A. [Ed], Cerebral Mechanisms in Behavior, 112-136, New York: John Wiley & Sons, 1951.

[33] Cummins, F. & Port, R. F., Rhythmic commonalities between hand gestures and speech, in Proc. Eighteenth Annual Conference of the Cognitive Science Society, 415–419, Mahwa, NJ: Lawrence Erlbaum Associates, 1996.

[34] Rossini, N., "Gesture in the brain: a multi-tasking experiment", in Zlatev, J., Johansson Falck, M., Lundmark, C. and Andrén M. [Eds], Studies in Language and Cognition, 436-453, Cambridge Scholars Publishing: Cambridge, 2009.