

Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues

Maciej Karpiński,* Ewa Jarmołowicz-Nowikow,** and Zofia Malisz***

***Center for Speech and Language Processing AMU

**Institute of Linguistics AMU

***School of English AMU, Poland

maciej.karpinski@amu.edu.pl

ewa@jarmolowicz.art.pl

zmalisz@ifa.amu.edu.pl

ABSTRACT

In the present paper, some preliminary findings regarding the structure of multimodal utterances in Polish task-oriented dialogues are discussed. The analysed material included six dialogue sessions. Instruction giver's utterances were transcribed phonemically, segmented into syllables, annotated for prominences and divided into minor and major intonational phrases. Independently, the corresponding video data were tagged for hand gestures. A probabilistic model of the gesture phrase was proposed. The commonly agreed coincidence of particular units in the streams of speech and gestures was tested for gestural and intonational phrases as well as for strokes and strong prosodic prominences. It was shown that the widely accepted idea of speech-gesture temporal alignment still needs more detailed research, especially in the context of dialogue interaction and in terms of units of analysis.

1. Introduction

The aim of the present work is to describe selected features of multimodal utterances in Polish task-oriented dialogues. The analysis is focused on prosodic properties of utterances and on hand gestures. Special attention is paid to the temporal coordination of prosodic and gestural units. In this type of analysis, a reference to the meaning of utterances is frequently necessary and therefore it still cannot be done fully automatically. The under-defined nature of gestural and prosodic units as well as their internal structure also rise many controversies. Moreover, there are many approaches to the studies of timing relationship between gesture and speech. In order to clarify our position, some relevant theoretical issues are briefly reviewed prior to the description of the study.

2. Speech and gesture: Common sources and mutual relations

The notion of a "common source" of speech and gesture may be interpreted in many ways. The interpretations are interconnected and should all be considered in the development of multimodal communication models.

First of all, it has been hypothesized that a *biological and evolutionary source* of both behaviours in the neuronal structure of the brain is possible to be found. The location of mirror neurons in the Broca's area may imply that the actual linguistic specialisation of the area evolved from premotor areas controlling gesture. On these grounds, Arbib [2005], Corballis [2002] and others suggested that speech developed from some form of communicative gesture. McNeill *et al.* [2005] use the same neuroscientific evidence to support an alternative hypothesis saying that speech and gesture evolved simultaneously without a "gesture-first" step (*ibid.* for discussion). Secondly, both behaviours share *similar motor control mechanisms* [Gentilucci & Dalla Volta 2007]. Speech and gesture use movements of the body, be it the vocal tract or the hands, to signal meaning. The movements have a lot in common not only in terms of neuronal correlates but also in terms of their dynamics. Attempts to develop a unified model of speech and gesture motor coordination have been recently called for [Iverson & Thelen 1999; Treffner & Peter 2002].

Moreover, from an essentially linguistic point of view, both modalities exist in everyday communication alongside each other and serve a similar semantic and pragmatic function of conveying ideas. So it can be said that the *cognitive common source* is an idea or ideas that are expressed multimodally in an utterance (McNeill and Duncan's "growth point" [2000]). How the two modalities, speech and gesture, support and relate to each other in fulfilling the fundamental communicative function is of importance to the present study.

The strength of the relation between gesture and speech is a matter of debate [Loehr 2004; Kendon 2007]. It is agreed that speech can function without gesture in an efficient way and that presence of speech-related gesture, as opposed to sign language or emblems, is by definition dependent on the presence of speech. Speech-accompanying gestures are important both for the speaker and the hearer. Apart from displaying attitudes and emotions, they add and enhance information primarily conveyed by speech.

Studies show that people gesture regularly while on the phone or when nobody is watching. However, they do not gesture while speaking into a tape recorder [Bavelas *et al.* 1992, 1995]. Blind speakers, also those born blind, gesture when in conversation with other blind people [Iverson & Goldin-Meadow 1998]. Gesture helps healthy speakers retrieve words from memory, hence they often coincide with hesitations in speech [Esposito & Marinaro 2007]. And of course, gestures are the modality of choice when speech and/or hearing fails due to impairment or circumstance. All these instances show that gesture is an instant and efficient partner when speech has problems. Naturally, there are also advantages of using gestures in a non-problematic spontaneous interaction. Although speakers rarely control gestures consciously, they cannot help using them in regular speech interaction and are aware of the results of the gestures. It can be in fact argued that since speech and gesture aim at a specific communicative goal, they reflect, in partnership, a common intention of the speaker. Conversely, since their outcomes are also perceived by the listener, they contribute, together, to the listener's understanding of the message. Gestures naturally align with speech when speech and gesture share underlying meanings in discourse.

Because spontaneous gesture is not codified, unlike speech or sign language, it helps convey and form new ideas, especially those that can benefit from spatial representation. Gestures actually offer an alternative format to represent thoughts, not

necessarily mutually exclusive with speech. Because, as Goldin-Meadow [Goldin-Meadow 1999; cf. McNeill 2007] points out, speech is “categorical, linear and discrete” and gesture is “analog and mimetic” (shapes, sizes etc.) communication and thinking is enhanced by the use of the two different representational devices. She shows that the enhancement is apparent in gesture-language mismatch, when both modalities convey different messages at the same time or when two ideas are actually being formed at the same time. Alternatively, Esposito notes that the disparate realisations may also reflect a unified planning process that differs only in the implementation used naturally by the two modalities [Esposito 2007:53].

3. Units in speech and gesture analysis

Units of gesture and speech may be defined on various levels of analysis, starting with physical, easily measurable phenomena, ending with “mental entities” (like dialogue acts) that are barely “reflected” in actual utterances. For both modalities, continuous arrays of phenomena are divided into segments with a degree of arbitrariness. It results in a range of problems faced in the process of gesture and speech segmentation. In everyday communication, a substantial proportion of incomplete, distorted utterances occur. As opposed to those compliant with a given model, they are frequently referred to as “ill-formed”. Even the determination of boundaries of the analysed units poses serious problems (e.g., [Ladd 1996:235]). The number of various models and approaches to the internal structure of prosodic and gestural units may reflect the complexity of the phenomena under study.

3.1. Units in the analysis of gestures

The literature on the *gesture phrase* is not as extensive as the one on intonation. Kendon [1972] was the first to introduce phrasal analysis of gesturing and to describe structure of gesturing in terms of *gesture units*, *gesture phrases*, and *gesture phases*. Kendon’s model of the *gesture phrase* was extended by other researchers. Kita [1990] added *pre-* and *poststroke hold phases*, Duncan extended the system with the notion of *stroke hold phases* for motionless *strokes* (quoted in [McNeill 2005]) and Kipp contributed with the *recoil phase* [Kipp 2004].

Kendon’s model of the *gesture phrase* (in the literature *gesture phrase* is used as a synonym of *gesture*) consists of the *preparation*, *stroke* and *recovery phase* (McNeill names the last one *retraction phase*). The *stroke* (the only obligatory phase) is the “most meaningful part” of a gesture. It may be repetitive and multi-segmented, so a few smaller movements may be involved in one *stroke phase* [Kendon 2005; McNeill 2007; Kipp 2004]. The *stroke* is the center of the *gesture phrase*, remaining phases are organized around it. The phase that leads to *stroke* is called the *preparation phase*. *Preparation* begins when the hand moves away from the resting position and ends just when the *stroke* begins. The *recovery phase* may be distinguished when the hand moves back to the resting position. It may happen that just before the *stroke* begins, the hand is held still for a moment. This position of the hand in gesture space is called a *pre-stroke hold phase*. A *poststroke hold phase* occurs when a corresponding situation is taking place at the end of the *stroke*. *Stroke holds* are *strokes* in the sense of meaning

and effort but they do not require movements of the hand. A *recoil phase* is a small hand movement after a forceful *stroke* when the hand lashes back from the *stroke* end position.

Gut *et al.* points out, that there are two different kinds of *gesture phases* that form a *gesture phrase* proposed by researchers: *function-oriented gestural phases* with functional interpretation (described below) and *form-oriented phases* (*source, trajectory, target*) describing only the form of a gesture [Gut *et al.* 2003].

3.2. Units in the analysis of prosody

Intonation units introduced by various researchers sometimes share only basic characteristics (for more detailed review of these issues, see [Cruttenden 1986] or [Fox 2000]). They usually have an internal structure which involves some perceptually prominent events and they can be extracted from the flow of speech on the basis of some systematic criteria. They are also described as showing “melodic completeness”. But the details of defining approaches strongly vary from one author to another. While the intonation unit is to be regarded as a linguistic entity, its definition may not involve the acoustic properties only. The completeness of its semantic and syntactic structure should be considered as well [Demenko & Jassem 1997]. There are also controversies related to the internal structure of intonation units. Traditional models, especially those belonging to the British school, seem to lean towards the idea of one central prominence around which the intonation unit is built (e.g., [Palmer 1922; Kingdon 1958; O’Connor & Arnold 1973], however opposite views exist (for a brief review, see [Baranowska *et al.* 2003]). It is also suggested that intonational segmentation may involve a hierarchy of more than one level. Beckmann & Pierrehumbert [1986] introduced the notion of an *intermediate phrase*, a constituent of the *intonational phrase*. *Intermediate phrases* are delimited on the basis of perceived degree of disjuncture and of pitch contours.

The properties and role of prosodically prominent syllables around which intonational units are built are also a matter of debate. They definitely contribute to the rhythmic structure of utterances. Prosodic prominence can be achieved using various acoustic dimensions (length, loudness, pitch) and these means may be language-specific. Objective judgment of the degree or level of prominence may prove extremely complex due to the fact that many bottom-up and top-down factors are involved, and the process of perception refers to a series of events that spread in time.

4. Temporal alignment of speech and gesture: Classical hypotheses and recent developments

Observations carried out by researchers show a lot of evidence that gestures and speech are tightly intertwined. We review two connected hypotheses first. According to McNeill there are three “rules” concerning speech and gesture synchronisation: semantic synchrony rule (speech and gesture present the same meaning at the same time), pragmatic synchrony rule (gesture and speech have the same pragmatic function) and phonological synchrony rule (gesture phrases coincide with tone units) [McNeill 1995]. Some implications of the last rule are examined in the present study.

The phonological synchrony rule is mainly based on Kendon’s observations [Kendon 2005]. According to Kendon, the synchrony on the phonological level concerns the

stroke of the gesture phrase and the tonic syllable of the co-occurrent tone unit. The rule says that the *stroke* of the gesture phrase is always completed either before or at the same time as the tonic syllable of the co-occurrent tone unit. It was also noted that the preparation phase precedes the tone unit with which it is associated. Unfortunately, within the phonological synchrony rule, it is not possible to define the meaning of “precedence” in terms of absolute time units. McNeill comments on Kendon’s observations saying that gestures both anticipate and synchronize with speech, bearing in mind that the anticipation and synchronisation refer to different phases of the gesture [McNeill 1992].

The timing relations between gesture and prominence investigated in the literature were usually realised by pitch accented syllables and different types of gesture effort peaks. McClave [1991] (three speakers) and Loehr [2004] (four speakers, selected parts) set out to verify Bolinger’s observation [1983, 1986] that gestures, in their direction of movement, follow pitch contours up and down, in parallel. The phenomenon occurred occasionally in McClave’s and Loehr’s data, however they found no significant correlation. McClave also showed that in the case of a complex gesture (*cf.* our results) where several movements appear in succession, the gestures are “compressed” and fronted to all finish before a stressed syllable. Additionally, gesture phrases in Loehr’s data typically slightly preceded the corresponding intermediate intonational phrases. He, as well as Jannedy and Mendoza-Denton [2005] (one speaker), Yasinnik *et al.* [2004] (three speakers) in analyses of the alignment of pitch accents and gesture effort peaks (also called “apices” or “hits”) clearly demonstrated that apexes of gestural movement are aligned with pitch accents. There were almost no apexes that did not coincide with a pitch accent in Jannedy and Mendoza-Denton’s data [2005:232]. It has to be noted that all of the above studies were based on monologues in English.

5. The co-occurrence of gesture and speech phenomena in the DiaGest Corpus

5.1. The material under study

The material under study comes from six task-oriented dialogues based on an “origami task”. The task involves a reconstruction of a paper folding invisible to the Instruction Follower who relies only on the Instruction Giver’s guidance [Jarmołowicz *et al.* 2007]. Six subject pairs were audiovisually recorded in a sound-proof studio. The audio recordings were transcribed both orthographically and phonemically in SAMPA, segmented into syllables, divided into intonational phrases, and labelled for strong and weaker prosodic prominences (*strong PP* and *weaker PP*, respectively). Phrasing was carried out according to guidelines by [Karpiński 2006], extended with the concept of two-tier phrase segmentation introduced in [Wagner 2008]. Major and minor intonational phrases (*Major IPs* and *Minor IPs*, respectively) were tagged on separate tiers and their well-formedness was arbitrarily judged against some basic criteria (no interruptions, no hesitations, no exaggerated lengthening). Prominences were tagged according to the guidelines of the *RaP* system [Dilley 2005]. Two levels of prominence were tagged on a separate tier solely on the basis of perception (careful listening) with no reference to higher domain linguistic knowledge. Speech signal segmentation and annotation was

Table 1. The statistical profile of the material under study

Units	Items tagged	Average per speaker
Syllable – Instruction Giver	4866	811
Syllable – Instruction Follower	1441	240.2
Strong prominence	755	125.8
Weaker prominence	1031	171.8
G-Phase	713	118.8
G-Phrase	223	37.2
Minor IP (well-formed)	905 (708)	150 (118.5)
Major IP (well-formed)	773 (539)	128.8 (89.8)

carried out using Praat. Resulting files were imported into ELAN (software by MPI) and integrated with independently prepared gestural annotation. The gestural annotation included a number of tiers, confessed to hand movements as well as to the segmentation of the gestural flow into gestural phases (*G-Phases*) and phrases (*G-Phrases*). It was based mostly on McNeill's model of the *G-Phrase*.

5.2. Gestural Phrases and Intonational Phrases

G-Phrases and *IPs*, as understood in this study, bear some structural resemblance. Both are built around a central prominent event (the accent in speech and the stroke in the gestural modality) and, in both cases, this event is the only obligatory component of the unit. In this section, the temporal co-occurrence of *G-Phrases* and *IPs* is analysed.

In the entire analysed material, 223 *G-Phrases* consisting of 713 *G-Phases* were tagged. In the utterances produced by instruction givers, 905 *Minor* and 773 *Major IPs* were found. Fluent, coherent and grammatically acceptable *IPs* were marked as “well-formed” (*WF IPs*) while others were generally categorised as “ill-formed” (*IF IPs*) and tagged for the type of defect.

In Figure 1, the frequencies of *G-Phases* that occur in the studied material are represented. While the number of *strokes* is equal to the number of analysed *G-Phrases*, the optional components of the *G-Phrase* clearly vary in their frequencies. *Pre-stroke hold* is the rarest category while *preparation* occurs quite frequently.

The presented data may be regarded as a simple probabilistic model of the *G-Phrase* in Polish task-oriented dialogues which describes the *G-Phrase* in terms of its more or less expected components, i.e. *G-Phases*.

In order to analyse the co-occurrence tendencies for *G-Phrases* and *IPs*, a number of queries were carried out using an advanced search tool provided by ELAN. Figure 2 illustrates how many cases of initial boundary overlaps, inclusions (*IP* is within the temporal limits of the *G-Phrase*) and final overlaps (where an *IP* overlaps with the final boundary of the *G-Phrase* in focus) occurred in the analysed material. The total number of occurrences is higher than the number of the *G-Phrases* due to the fact that some *G-Phrases* overlapped with more than one *Major IP*. The results certainly confirm the claim of the co-occurrence of corresponding, semantically bound gestural and intonational units, however the exact synchrony seems to be disputable.

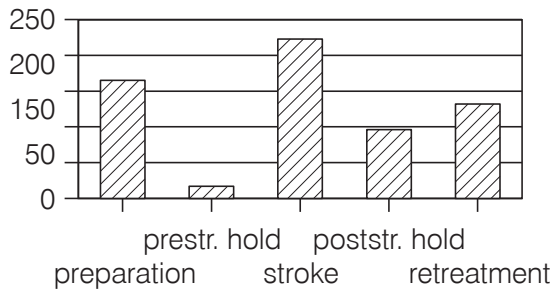


Figure 1. The quantities of respective G-Phases found in the studied material.

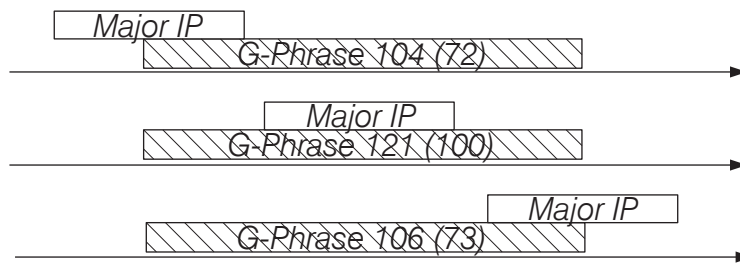


Figure 2. The number of G-Phases overlapping any Major IP (in brackets, the numbers of WF Major IPs meeting respective conditions are given).

5.3. Strokes and prominences

An attempt was made to study the details of the co-occurrence of *strokes* and prominences, similarly to *G-Phrases* and intonational phrases in 5.2. Particularly to see whether the phonological synchrony rule (henceforth *PSR*) [McNeill 1995] holds for spontaneous, interactive communication.

At the first stage, *strokes* and *prosodic prominences* (henceforth *PPs*; defined as *beats* in the *RaP* system) were analysed for number of overlaps. Out of 223 *strokes*, 96% overlapped with any *PP*, but only 75% overlapped with *strong PPs*. Among overlaps between the *stroke* and the *strong PP*, initial overlaps (only the initial boundary of a given *stroke* is within the temporal limits of the respective *strong PP*) accounted for 26% of cases and final overlaps (only the final boundary of a given *stroke* lies within the respective *strong PP*) accounted for 38% of cases, while the cases of inclusion (*strong PP* within the limits of a *stroke*) were noted in 46% of the analysed overlap cases. This preliminary result may suggest that the *PSR* is frequently violated because only final overlaps ensure that the final boundary of a *stroke* precedes the final boundary of the respective *strong PP*. (Please note that in a number of cases there was no overlap at all between the units in question, and the *PSR* was satisfied because the entire *strong PP* corresponding to a given *stroke* followed it with no overlap). A more detailed qualitative analysis was carried out and the following categories of *PSR* violation were distinguished:

1. Long *stroke* with a continuous gestural excursion.
2. Long and complex *stroke* may exceed the final boundary of the corresponding *strong PP* syllable. This may happen especially with hand movements repeated intentionally for more clarity or expression. The first realisation always meets the *PSR*, but the following repetitions fall behind the final boundary of the corresponding *strong PP* syllable.
3. Sometimes the first movement of the *stroke* seems to be realised fully consciously while the following ones are just automatic repetitions and they “echo” or “copy” the first one in a rhythmic manner that resembles *batons* (they also seems to be physically “weaker”, less clear versions of the first one).
4. Quick and energetic *strokes* tend to meet the rule. However, they are frequently followed by minute movements occurring as a result of inertia labelled as a separate category of “recoil” by Kipp [2004]. This final stage of the *stroke* may exceed the boundary of the corresponding *strong PP* syllable.
5. In a phrase with an initial strong prominence, following a pause, if the *stroke* starts almost simultaneously with the initial syllable of the phrase, it usually continues after its final boundary violating the *PSR*.
6. With ill-formed *IPs* and *G-Phrases*, the rule is often not valid. The focus may not be realised at all in the case of an incomplete *IP*, or the *stroke* may be difficult to notice or interpret when it is realised in a distorted or incomplete way. Also, the *PSR* often fails when disrupted by gestural or speech hesitations.

While the above categories account for the majority of violations, some cases cannot be explained in the above terms. As the *PSR* seems to apply mostly to narration, one may suspect that the interactivity of dialogue introduces other factors that may influence the timing of speech and gesture.

A preliminary analysis of the onset timing for the *stroke* and the corresponding *strong PP* shows that the distance between the beginning of the *stroke* and the beginning of the prominent syllable is shorter than 100 ms only for ca. 5% of the cases and shorter than 200 ms for ca. 40%.

6. Conclusion and further studies

The boundaries of *G-Phrases* and *IPs* do not seem to synchronise very precisely. Still, in most cases, there is at least an overlap of *G-Phrases* and respective *Major IPs*. This is, however, not surprising, since gestures co-occur with speech and consequently, the probability that they will at least overlap with the neighbouring *IP* is very high. Nevertheless, there is a clear symmetry in the number of initial and final overlaps. This finding seems to reveal a general “centering tendency” for the semantically related *G-Phrases* and *Major IPs* and confirm their temporal co-occurrence.

Although the *PSR* seems to hold for the majority of analysed cases, *strokes* and respective *strong PPs* violate the *PSR* relatively frequently, mostly in situations listed in 5.3. The exceptions can be justified by the characteristics of spontaneous dialogue interaction as well as by the constraints of the dialogue task. The task requires extensive gesturing and complex sequences of hand movements are often involved that are different from those in everyday conversations.

In order to clarify some of the findings, and extend the scope of analysis, it is planned to add point-centred tagging to the present labelling system, i.e. tagging *P-Centres* in prominent syllables and peak effort points of gestural trajectories. Further studies will also involve additional intonation labelling, mostly in terms of f_0 peaks.

BIBLIOGRAPHY

- Arbib, M. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28: 105–124.
- Baranowska, E., Francuzik, K., Karpiński, M., Kleśta, J. 2003. Identification of Nuclear Melody Placement in Polish Read Texts [In:] A. Mettouchi & G. Ferre (Eds.) *Interfaces Prosodiques*, Nantes.
- Bavelas, J., Chovil, N., Lawrie, D., Wade, W. 1992. Interactive gestures. *Discourse Processes* 15, pp. 469–489.
- Bavelas, J., Chovil, N. 1995. Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21(4), pp. 394–405.
- Beckman, M. E. & Pierrehumbert, J. B. 1986. Intonational structure in Japanese and English. *Phonology Yearbook*, 3, pp. 255–309.
- Bolinger, D. 1983. Intonation and gesture. *American Speech*, 58(2), pp.156–174.
- Bolinger, D. 1986. *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.
- Corballis, M. C. 2002. *From hand to mouth: the origins of language*. Princeton, NJ: Princeton University Press.
- Cruttenden, A. 1996. *Intonation*. Cambridge: CUP.
- Dilley, L. & Brown, M. 2005. *The RaP Labeling System*, v. 1.0, ms. <http://faculty.psy.ohiostate.edu/pitt/dilley/rapsystem.htm>
- Esposito, A. & Marinaro, M., 2007. What pauses can tell us about speech and gesture partnership. In: Esposito, A., Bratanić, M., Keller, E., Marinaro, M. (Eds.) *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*. Amsterdam: IOS Press.
- Esposito, A., Bratanić, M., Keller, E., Marinaro, M., 2007. *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*. Amsterdam: IOS Press.
- Fox, A. 2000. *Prosodic Features and Prosodic Structure*. Oxford: OUP.
- Gentilucci, M. & Dalla Volta, R. 2007. The motor system and the relationship between speech and gesture. *Gesture* 7:2, 140, pp. 159–177.
- Goldin-Meadow, S. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 11 (3), pp. 419–429.
- Gut, U., Looks, K., Thies, A., Gibbon, D. 2003. CoGesT: Conversational Gesture Transcription System. Version 1.0. Technical report. Bielefeld University.
- Iverson, J., Goldin-Meadow, S. 1998. Why people gesture as they speak? *Nature* 396, 228.
- Iverson, J., Thelen, E. 1999. Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6, pp. 19–30.
- Jannedy, S., Mendoza-Denton, N. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, 3, pp. 199–244.
- Karpiński, M. 2006. *Struktura i intonacja polskiego dialogu zadaniowego*. Poznań: Wydawnictwo Naukowe UAM.

- Kendon, A. 2005. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, A. 2007. Some topics in gesture studies. In: Esposito, A., Bratanić, M., Keller, E., Marinaro, M., *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue. 2007*. Amsterdam: IOS Press.
- Kingon, R. 1958. *The Groundwork of English Intonation*. London: Longman.
- Kipp, M. 2004. *Gesture Generation by Imitation. From Human Behavior to Computer Character Animation*. Boca Raton: Dissertation.com.
- Kita, S., Gijn van, I. & Hulst van der, H. 1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth, I., Fröhlich, M. (Eds.) *Gesture and Sign Language in Human-Computer Interaction*. Berlin: Springer Verlag, pp. 23–35.
- Loehr, D. P. 2004. *Gesture and intonation*. An unpublished PhD dissertation, Georgetown University, Washington DC.
- McClave, E. 1991. Intonation and gesture. Doctoral Dissertation, Georgetown University, Washington DC.
- McNeill, D. 1995. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, D. 2007. *Gesture and Thought*. Chicago: University of Chicago Press.
- McNeill, D., Duncan, S. 2000. Growth points in thinking for speaking. In: McNeill, D. (Ed.) *Language and gesture*. Cambridge: CUP, pp. 141–161.
- McNeill, D., Bertenthal, B., Cole, J., Gallagher, S. 2005. Gesture-first, but no gestures? Commentary on Michael Arbib “From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28, pp. 105–124.
- Palmer, H. E. 1922. *English Intonation with Systematic Exercises*. Cambridge: Heffer.
- Treffner, P., Peter, M. 2002. Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, 21, pp. 641–697.
- Wagner, A. 2008. A comprehensive model of intonation for application in speech synthesis. Doctoral Dissertation, Adam Mickiewicz University, Poznań.
- Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S. 2004. The timing of speech-accompanying gestures with respect to prosody. *Acoustical Society of America Journal*, 115 (5), pp. 2397–2397.