Gesture annotation
and what can be squized out of it
for the analysis of alignment

Agnieszka Czoska & Maciej Karpiński
Center for Speech & Language Processing
AMU Poznań

Views to entrainment/alignment

Pickering & Garrod (2004) ascribe linguistic alignment largely to an automatic priming of lexical, syntactic, or semantic representations and a percolation of activation between adjacent representational levels. Alignment occurs when the two interlocutors employ equivalent representations at different levels, and arises from an automatic priming mechanism" (but don't forget about self-alignment!)

Brennan & Clark (1996) claim that speakers strategically design utterances for an addressee and thereby prefer previously used (grounded) constructions.

Mol et al. (2009): "[...] the mimicry of iconic gestures is an instance of convergence of linguistic behavior, also known as alignment, rather than a superficial imitation of physical behavior."

Conditions to meet, assumptions to make & decisions to take

- **Understandings and views of „alignment", e.g.**
  - Narrow vs. Wide (only convergence, similarity or more?)
  - Deep vs. Shallow, Automatic vs. Conscious
  - Immediate vs. Long-run, gradual, etc. (static vs. procesual)
- **Where to look for similarity/priming?**
  - **B**etween which particular phenomena (and which of their properties)?

Some of the available systems

NeuroGes (Lausberg et al.) - multi-level

Mumin (Allwood et al.)

CoGest (Gut et al.)

Form (Martell) – detailed but time-consuming

Anvil (Kipp et al. , associated with Anvil software)

DiaGest (Jarmołowicz-Nowikow et al., used with Elan)

An example of a multi-tier annotation system (DiaGest in ELAN)

Conditions to meet, assumptions to make & decisions to take

- **Limited amount of recordings**
  - Limited number of cases/realisations of assummed categories or examples of gestural feature
  - Representativeness
- **Limited view/visibility**
- **Presumably culture-specific aspects/meanings to be preserved**
  - cannot reject certain phenomena as irrelevant without consulting culture experts
  - low-level annotation may be safer when it comes to catching
- **Side-effects of the task characteristics** (some categories of „gestural events" more probable, apparent alignment may result)
- **Limited time / power for annotation** (cannot afford to annotate too much "just in case")

The choice of phenomena to be coded

- Physical description of motion & shape
    - Directions, peaks, velocities
    - Gestural phrase and its phases (more arbitrary...)
- Gestural „categories"
    - e.g., from Kendonian... continuum ;-)
- Gestural features
    - e.g., mode of representation, handedness

Potentially relevant features

Bergmann & Kopp (2012) on gestural form convergence:

„We found evidence for gestural alignment but not all the features of co-speech gestures are subject to this effect

Similarity of gestures measured in terms of a set of „simple" features:

(1) HANDEDNESS

(2) HANDSHAPE (ASL-based coding)

(3) PALM- AND FINGER ORIENTATION

(4) WRIST MOVEMENT TYPE (static, linear, or curved + sequences).

(5) REPRESENTATION TECHNIQUE

Proposed solution

- Multi-tier
  - Hand movement tiers
  - Head movement tier(s)
  - Body position tier(s)
  - Gaze tier (?)
- Annotation on the level of physical features necessary to minimise the „cultural bias"
- Annotation system – not only as a set of symbols but **also a procedure** (algorithm?)

Head movement annotation

Table 1: Head gesture Inventory.

Label	Description	Axis
Nod	Rotation down-up	Pitch
Jerk	'Inverted nod', head upwards	Pitch
Tilt	'Sideways nod'	Roll
Shake	Rotation left-right horizontally	Yaw
Pro	Pushing the head forward	Z
Retr	Pulling the head back	Z
Turn	Rotation left OR right	Yaw
Bobble	Shaking by tilting left-right	Roll
Slide	Sideways movement(no rotation)	X
Shift	Repeated slides left-right	X
Waggle	Irregular connected movement	

+ potential kinematic specifiers (?), e.g. energy, direction, periodicity,

(Kousidis et al. 2011)

●

A possible approach
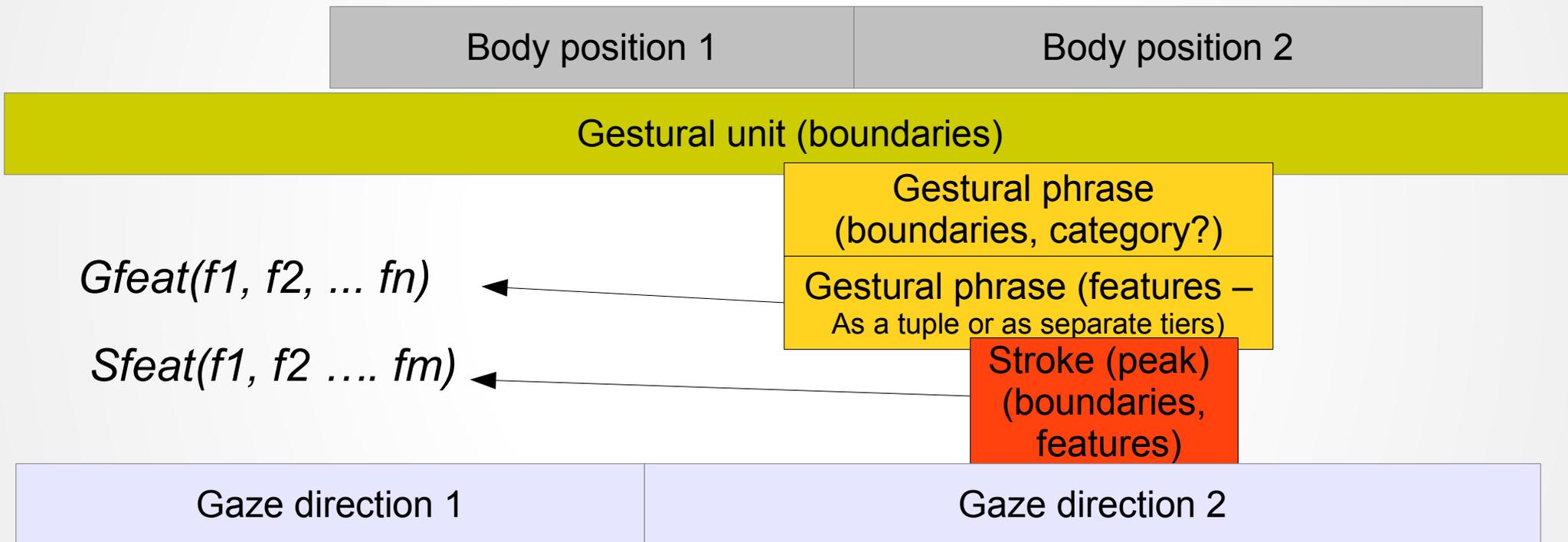
Some references

Brennan, S. E., Galati, A., K. Kuhlen, A. Two Minds, One Dialog: Coordinating Speaking and Understanding. The Psychology of Learning and Motivation

Mol, L., E. Krahmer, A. Maes and M. Swerts (2012). Adaptation in gesture: Converging hands or converging minds? Journal of Memory and Language, 66, pp. 249-264.

Pickering, M. J. & Garrod, S. 2004. Toward a mechanistic psychology of dialogue, Behavioral and Brain Sciences, vol. 27, pp. 169–226.

Pickering, M. J. & Garrod, S. 2006. Alignment as the Basis for Successful Communication. Research on Language and Computation.

Włodarczak, M., Buschmeier, H., Malisz, Z., Kopp, S., & Wagner, P. (2012). Listener head gestures and verbal feedback expressions in a distraction task. In Proc. of the interdisciplinary workshop on feedback behaviours in dialogue

Some references

Allwood, J., Cerrato, L., Dybkær, L., Jokinen, K., Navarretta, C., & Paggio, P. (2004). *The **MUMIN** multimodal coding scheme*. Technical report available at www.cst.dk/mumin/stockholmws.html.

Gut, U., Looks, K., Thies, A., Trippel, T., Gibbon, D., 2002. **CoGesT**. Conversational Gesture Cranscription System. Version 1.0. Technical Report. University Bielefeld.

Kipp, M., Neff, M., and Albrecht, I. 2007. An Annotation Scheme for Conversational Gestures: How to economically capture timing and form. Journal on Language Resources and Evaluation - Special Issue on Multimodal Corpora, 41(3-4):325–339, December. (i.e., **ANVIL**)

Martell, C. 2002. **FORM**: An extensible, kinematically-based gesture annotation scheme," in *LREC Proceedings*, 2002, pp. 183–187.

Martell, C. 2005. FORM: An Experiment in the Annotation of the Kinematics of Gesture. Dissertation.

McNeill, D. 1995. Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press.

McNeill, D. 2007. Gesture and Thought. The University of Chicago Press, Chicago London

Thank you