Matching across Turns in Talk-in-Interaction: The Role of Prosody and Gesture

Jan Gorisch
05.02.2013

Supported by EU ‘Sound to Sense’ project [grant number MRTN-CT-2006-035561].
Department of Computer Science
Department of Human Communication Sciences
Overview

• Background
• Material and method
• Interactional analysis
• Prosodic analysis
• Gestural analysis
• Prosodic-gestural model
• Discussion and conclusions
Background
Alignment and non-alignment

Short turns, such as “uh huh”, “oh yeah”, “oh really”, perform two different actions, they ...

(1) align with the agenda of the prior speaker
   • Following which, the original speaker continues his agenda / project / action

(2) non-align with the agenda of the prior speaker
   • Following which, the original speaker does not continue on his own agenda.
     - speaker orients to the short turn in a way that demonstrates that it needs special treatment

[Stivers, 2008; Barth-Weingarten, 2011]
Background – prosodic patterns and social actions

• Specific prosody correlates with specific actions [e.g., Couper-Kuhlen, 2001]

• But: High prosodic variability even when same social action is performed [e.g., Geluykens, 1987]

• Indications that prosodic matching is used to perform specific social actions [e.g., Lerner, 2002]

• Conclusion
  – Prosodic analysis must be performed ACROSS turns and participants in order to account for the COORDINATION OF ACTIONS [Szczepesk Reed, 2012a]
  – Agenda alignment and non-alignment might be such actions
Research questions – prosody

• RQ1:
  – Are alignments performed with prosodic matches and are non-alignments performed with prosodic non-matches?

• How can prosodic similarity be measured objectively?
A speaker’s choice of pitch contour is explicable ...

- by reference to the pitch contour used by the speaker of the previous turn
- and this is determined by whether the current speaker aligns with the prior turn or not.
Background – gestures and social actions

• Specific gestures correlate with specific actions [e.g., Schegloff, 1987]

• Gesture matching correlates with specific actions [e.g., Selting, 2010]
Research questions – gesture

• RQ2a:
  – Are alignments and non-alignments performed with specific gestures?

• RQ2b:
  – Are alignments performed with gestural matches and are non-alignments performed with gestural non-matches?
Hypotheses – gesture

A: “Blah blah”

B: “Bluh bluh”

- H1: A specific gesture determines the social action.
  - E.g., a nod or head shake = alignment
- H2: A gestural match or non-match of the current gesture with the prior gesture determines the social action.
  - E.g., matching a prior nod or matching a prior head shake = alignment; non-matching = non-alignment
Research question – prosodic-gestural

• Analogy: prosody and gesture have a common origin [Kendon, 2004]
  – Tone units – prosodic features – discourse meaning
  – Gesture phrases – kinesic features – visible bodily actions

• RQ3:
  – Are the gestural domain and the prosodic domain correlated?
    • Is there an interaction between gestures and prosodic similarity in predicting the interactional category?
Material and method
Spontaneous research meetings from the AMI meeting corpus
http://corpus.amiproject.org

Meetings (EN2009b, EN2009c, EN2009d)
Native English speakers (MEE09, FEE083, MEE095 and FEE096)

Collection of 912 instance:

- short utterances; not more than a single intonational phrase (IP)
- produced by a “second” speaker
- either in the clear or in overlap
- preceded by a more extended turn from a first speaker
- immediately followed by another turn from that first speaker or by a new IP from the second speaker

Annotation of 8 major gesture types, e.g. nod, frown, raised eyebrows, ...
Example stretch of a meeting
Method – combination of CA and phonetic-prosodic-gestural analysis

• Conversation Analysis (CA)
  – Establish categories based on interactional evidence
  – Qualitative analysis by identifying participants’ orientation
  – Sequential organisation of talk
  – Collection of instances of social actions

• Instrumental analysis
  – Correlate categories with acoustic measures
  – Quantitative analysis using statistics
  – Sequential analysis of prosody (across speakers and turns)
  – Automatic extraction of prosodic features

Aim:
  • Correlation of interactional categories with prosodic similarity
  • Correlation of interactional categories with gesture types and with gestural copying behaviour
Interactional analysis
Outline

Interactional analysis – a coding exercise

- Description of the sequential organisation of aligning and non-aligning actions
- Classification of the 912 instances into interactional categories alignment and non-alignment
Questions, definitions and material

• What are the **sequential correlates** of the social actions alignment and non-alignment?

  – What is the **verbal and gestural organisation**?
  – Orthographic transcripts (without access to prosodic information)

Prior turn (talk/gesture)  

A: I say something

B: then I say something

time

Is the successive turn a **continuation** of the prior turn or not?

Then either of us says something

Successive turn (talk/gesture)

Current / target turn (talk/gesture)
# Continuation vs. non-continuation

<table>
<thead>
<tr>
<th>Verbal continuation</th>
<th>Verbal non-continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recycling of words or stretches of</td>
<td>No recycling of words</td>
</tr>
<tr>
<td>words</td>
<td></td>
</tr>
<tr>
<td>Continuous syntax and semantics</td>
<td>Disjoint syntax and semantics, sometimes resulting in a</td>
</tr>
<tr>
<td></td>
<td>logic break</td>
</tr>
<tr>
<td>Prior speaker occupies floor</td>
<td>Same speaker occupies floor</td>
</tr>
<tr>
<td>indicated by silences, hesitation,</td>
<td></td>
</tr>
<tr>
<td>false starts (dysfluency)</td>
<td></td>
</tr>
<tr>
<td>Turn “holding” or “trail-off”</td>
<td></td>
</tr>
<tr>
<td>conjunctons</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Successive laughter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gestural continuation</th>
<th>Gestural non-continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same gesture in the prior turn and</td>
<td>Different gesture in prior turn and successive turn</td>
</tr>
<tr>
<td>the successive turn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification task

• Purpose: database for subsequent analysis

• Are these just random observations or regularities?

• Classification results
  – 912 instances of adjacent turn pairs
    • 564 alignments
    • 348 non-alignments
  – 2 annotators
    • The author annotated all instances
    • A master student from HCS who was experienced in doing CA annotated 205 of 912 (Kappa = 0.64 -> substantial agreement)
Alignment

Extract 3: "drift correction" (B-03:27)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>Gest_A</td>
<td>but it loo* it looks like it was doing all the calibration and drift correction things that s*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2]</td>
<td>Gest_A</td>
<td>head shake</td>
<td>shrug</td>
<td>body movement</td>
<td>so that don't think there's anything...</td>
</tr>
<tr>
<td>[2]</td>
<td>Gest_B</td>
<td></td>
<td></td>
<td></td>
<td>nod and blink</td>
</tr>
<tr>
<td>[2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>yeah</td>
</tr>
<tr>
<td>[2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3]</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td>to add to the software (-) for that part ...</td>
</tr>
</tbody>
</table>

- Extended turn by A ending in line 2 with “so”, inbreath (“hh) and body movement
- B’s “yeah” is accompanied by a nod and a blink
- A continues successively with “so” and more talk on his agenda (reporting on the software development work)
- A does not hesitate subsequently and B does not interrupt (compete for the turn)
Extended turn by C ending with syntactic completion, inbreath ("hhh") and body movement

B’s "oh really" is accompanied by a nod and a blink first, then raised eyebrows

C first overlaps with continued talk on the prior agenda, then orients to the "oh really" by:

- Interrupting talk on the prior agenda (syntactic and semantic discontinuation)
- Nodding (gesture discontinuation)
- Treating the "oh really" as a turn that needs a response ("yeah mm-hmm")
- Not following the prior agenda

B acknowledges the response ("oh okay")
Discussion – interactional

• Agenda continuation and non-continuation are two different orientations to aligning vs. non-aligning turns.
  – They are consistently observable by an additional annotator.

• Findings support previous interactional phonetics research, proposing such or similar fundamental actions [e.g., Szczepak Reed, 2012a]
Prosodic analysis
Outline

Prosodic analysis

• Measuring prosodic similarity

• Correlation of interactional categories with similarity measure
Research questions

• RQ: Is pitch contour matching used for alignment, and is non-matching used for non-alignment?

• Hypothesis: *Aligning* turns will show higher prosodic similarity with the prior turn than *non-aligning* turns.
Sequential F0 comparison

\[ \text{sim}(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}} \]

\[ \sigma = 0.3 \]

[Cooke, 1993]
Accumulative quality score

- Pairwise comparison of F0-values $\rightarrow$ similarity matrix (left Fig.)
- Optimal accumulation of similarity scores (right Fig.) + normalisation for missing data $\rightarrow$ similarity metric $MD_{\text{norm}}$
Results ($\text{MD}_{\text{norm}}$)

- **Hypothesis**: *Aligning* turns will show higher prosodic similarity with the prior turn than *non-aligning* turns.

- **Significant difference** between interactional categories in prosodic similarity.

- **Mann Whitney U test**: difference between mean ranks is significant ($z=-10.057; p<0.001$)
  - “alignments” (n=533) mean rank: 504.23
  - “non-alignments” (n=338) mean rank: 328.41
Discussion – prosodic

• Results indicate that prosodic matching *is* a resource ...
  – to align with the agenda of the prior speaker
  – thereby to allow the prior speaker to continue on that agenda

• and non-matching is a resource
  – to non-align
  – and thereby to initiate a new action.
Gestural analysis
RQ2a:
- Are alignments and non-alignments performed with specific gestures?

RQ2b:
- Are alignments performed with gestural matches and are non-alignments performed with gestural non-matches?
H1: specific gestures correlate with interactional category

- Significant interaction between the interactional category and the current gesture:
  - $\chi^2(7, N=810) = 1.686E2$, $p<0.001$
  - Effect size: $\phi=0.46$

Alignments are more likely to be accompanied by nods
H2: gestural matching correlates with interactional category

- Association between the prior and current gesture is not significant
  - Alignments: $p=0.005$ ($\phi=0.407$; $V=0.154$)
  - Non-alignments: $p=0.113$ ($\phi=0.429$; $V=0.162$)

Alignments and non-alignments do not differ in terms of gestural matching
Discussion – gestural analysis

- It also suggests that the organisation of prosody and gesture are different
  - Possible explanation:
    - Gestures are easier to copy simultaneously;
    - Prosody is easier to copy in sequence
      - simultaneous prosodic matching or non-matching obscures each other.
Combined prosodic-gestural model
Outline

• Do gestures correlate with prosodic similarity?

• Combination into acoustic-gestural model
  – Comparison of gesture types and prosodic similarity
Prosodic-gestural model

- Significant interaction between gesture type and prosodic similarity
  - Kruskal-Wallis test: $\chi^2(7,N=766)=22.916$, $p=0.002$

- A nod is more likely than an eyebrows raise to co-occur with instances of high prosodic similarity
  - Mann-Whitney U test: raised eyebrows (n=35) x nod (n=319); $Z=3.243$, $p=0.001$ (Bonferroni adjusted)

H1: specific gestures and prosodic similarity interact in determining the interactional category
Discussion – prosodic-gestural model

• Weak interaction between prosodic similarity and gesture types (raised eyebrows vs. nod)

• Results do not support the analogy [Kendon, 2004]:
  – Tone units – prosodic features – discourse meaning
  – Gesture phrases – kinesic features – visible bodily actions
    • ... At least not for social actions of alignment and non-alignment

• The interactional category is predicted
  – Sometimes by specific gestures
  – Sometimes by prosodic similarity
  – In a few cases: gesture and prosodic similarity interact
Discussion and conclusions
Novel contributions

• Prosodic analysis
  – Sequential: prosodic matching
  – Automatic
  – Objective: based on signals

• Gestural analysis
  – Annotation of 3 AMI-meetings
  – Statistical analysis

• Interactional analysis
  – Relevant social actions
  – Inter-rater reliability
Answers to research questions

• Prosody:
  – Prosodic matching and non-matching are relevant resource for aligning and non-aligning actions.

• Gesture:
  – Individual gestures are more likely to cue the interactional category than gestural matching
    • Nods usually accompany alignments

• Prosodic-gestural model:
  – Nods are more likely than raised eyebrows to co-occur with an instance of prosodic matching
Interdisciplinary approach and Multimodality

• Interdisciplinary approach
  – CA+instrumental

• Multimodal
  – Analysis of verbal and non-verbal aspects of face-to-face meetings
Implications

• Modelling of prosody in ASR and speech synthesis (in dialogue systems)

• Child development of prosody [Wells, 2010]

• Atypical development, e.g. echolalia in autism [Local & Wootton, 1995]

• Teaching prosody to second language learners [Szczepk Reed, 2012b]

• The role of prosodic matching on empathy [Aziz-Zadeh et al., 2010]

• The origin of prosody and gesture
Thank you

j.gorisch@sheffield.ac.uk


end