



The  
University  
Of  
Sheffield.



# 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]

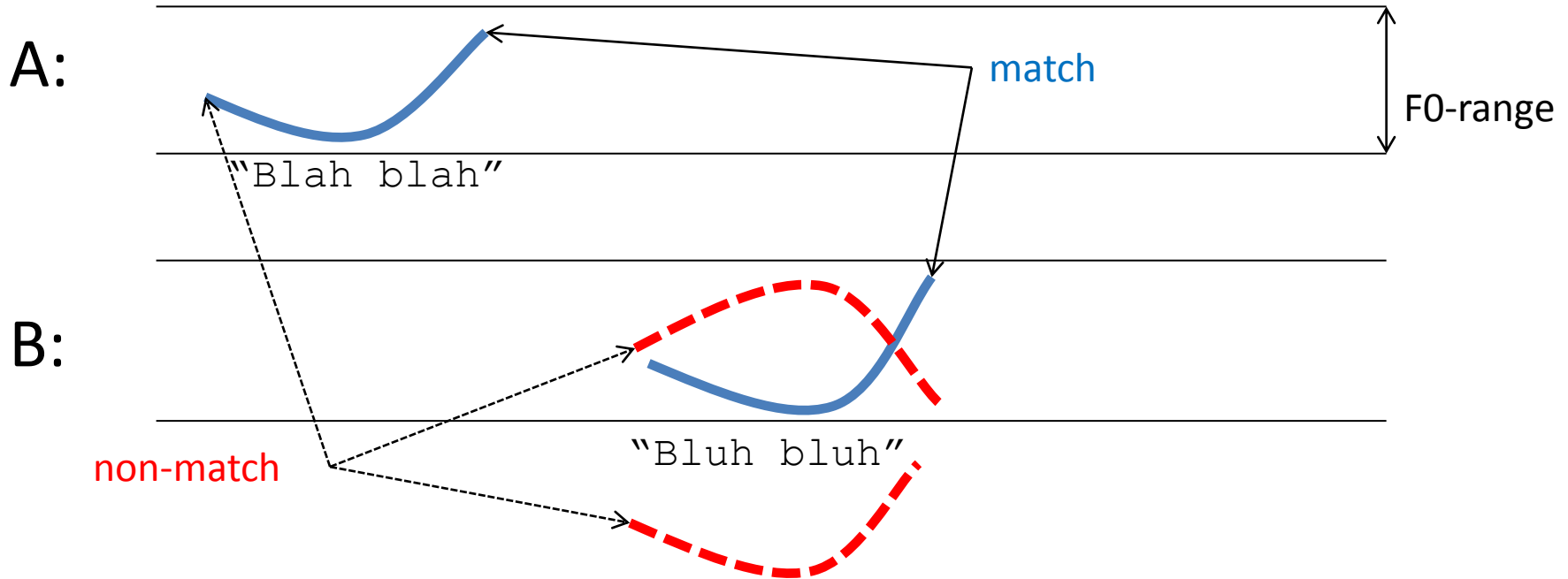
# Background – prosodic patterns and social actions

- 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** [Szczepk 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?

# Hypothesis – prosody



- 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*.

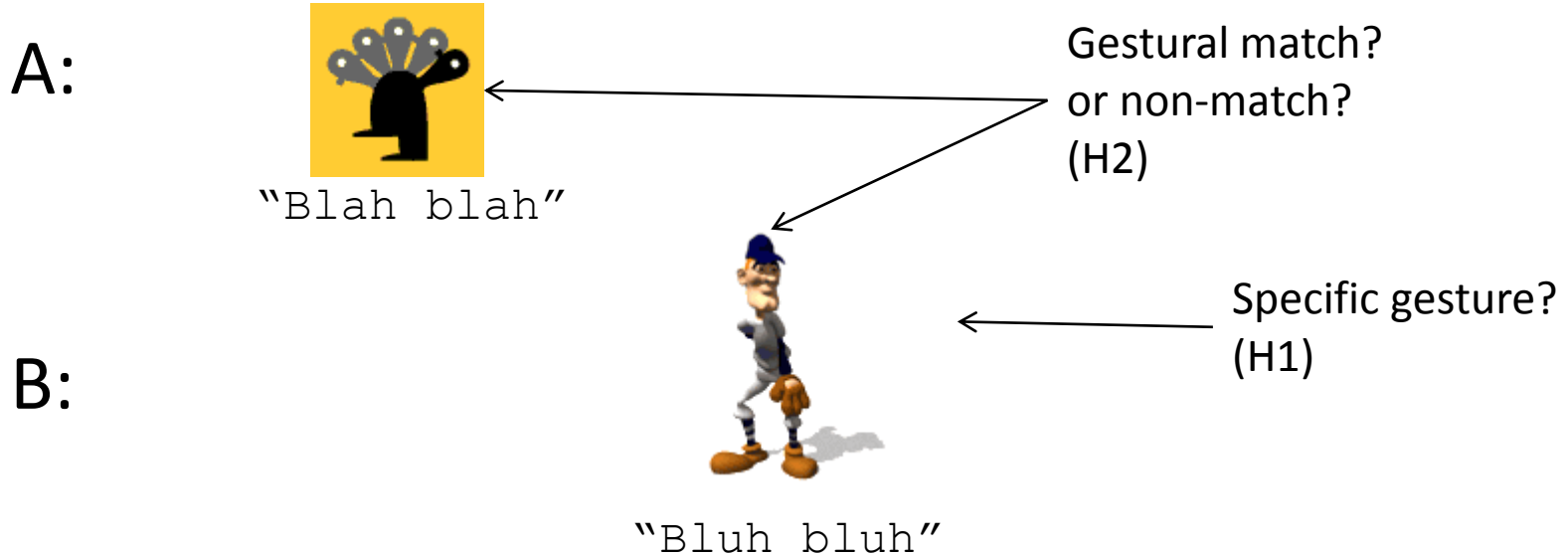
# Background – gestures and social actions

- 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**?

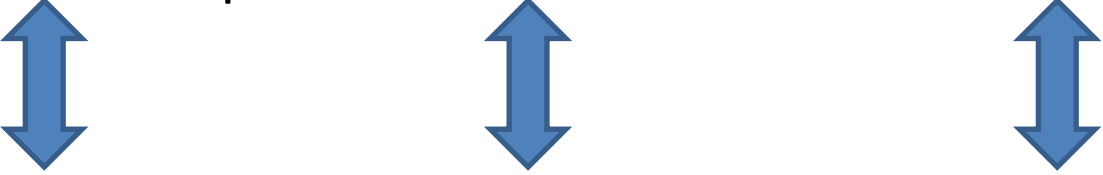
# Hypotheses – gesture



- 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



Equivalent!?
- 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



# Material

Individual video  
Headset mic.



Spontaneous research meetings from the AMI meeting corpus

<http://corpus.amiproject.org>

3 Meetings (EN2009b, EN2009c, EN2009d)

4 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

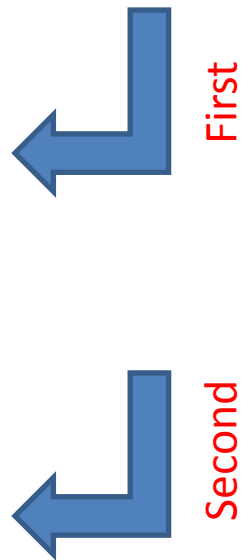


# Method – combination of CA and phonetic-prosodic-gestural analysis

- 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

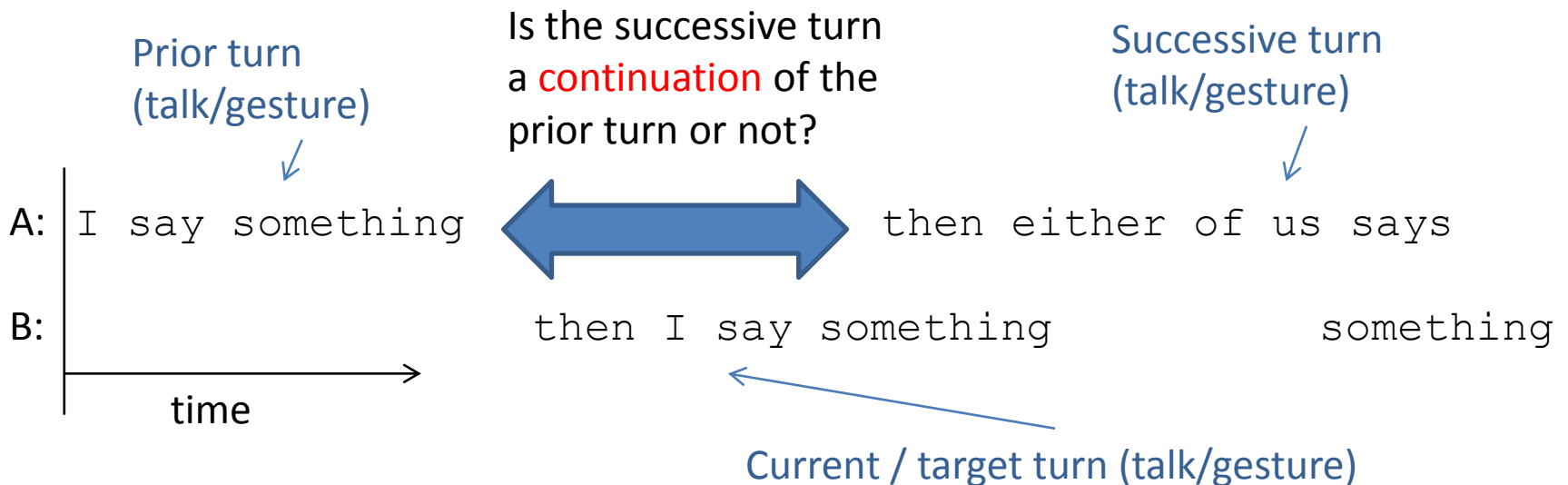
# Outline

## 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)



# Continuation vs. non-continuation

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

<b>Gestural continuation</b>	<b>Gestural non-continuation</b>
Same gesture in the prior turn and the successive turn	Different gesture in prior turn and successive turn

# 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)

[1]

Gest_A		<i>body movement</i>	<i>body movement</i>	
A	but it loo* it looks like it was doing all the calibration	and drift correction things	that s*	

[2]

Gest_A	<i>head shake</i>	<i>shrug</i>	<i>body movement</i>	
A	ellen was wanting on it so °hh		so that don't think there's an ything ...	
Gest_B			<i>nod and blink</i>	<i>nod</i>
B		yeah		
Gest_C		<i>nod</i>		<i>nod</i>
		(0.54)		

[3]

A	to add to the software (-) for that part ...
---	--

- 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)

# Non-alignment



Extract 8: "eye" (B-07:09)

[1]

Gest_C	<i>body movement</i>	<i>hand gesticulation</i>
C	ellen wanted something to look more like an eye (.) so that you knew that was the eye °h (0.44)	

[2]

Gest_A		<i>hand</i>
A		yeah so so we've got like uh a
Gest_B	<i>nod and blink</i>	<i>raised eyebrows nod and blink</i>
B	oh really	oh okay
Gest_C	<i>body movement</i>	<i>nod body movement body movement</i>
C	hh	there wa s like yeah mm-hmm (0.35) (0.4)

[3]

Gest A	<i>gesticulation</i>
A	big circle with a ...

- 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., Szczepek Reed, 2012a]

# Prosodic analysis

# Outline

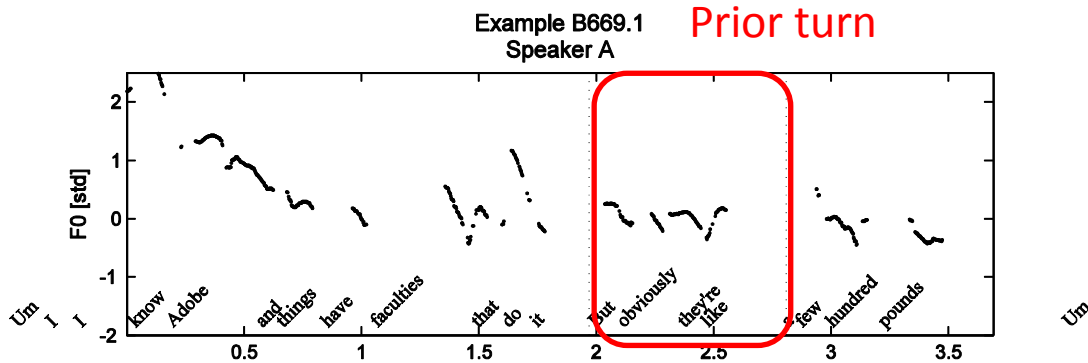
## Prosodic analysis

- Measuring prosodic similarity
- Correlation of interactional categories with similarity measure

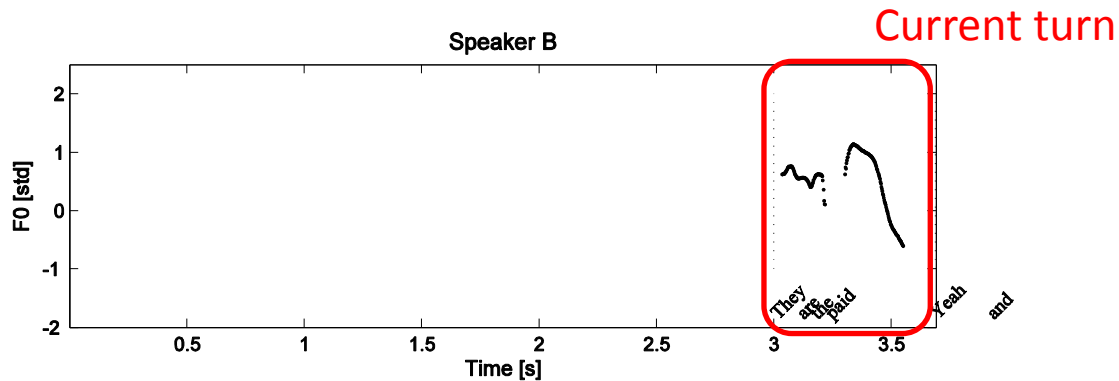
# Research questions

- 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

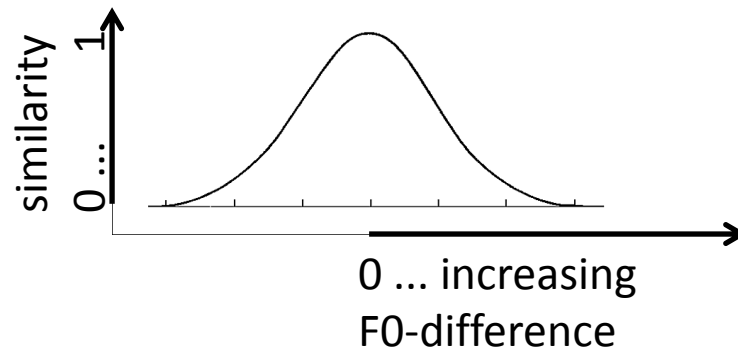


$$\text{sim}(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

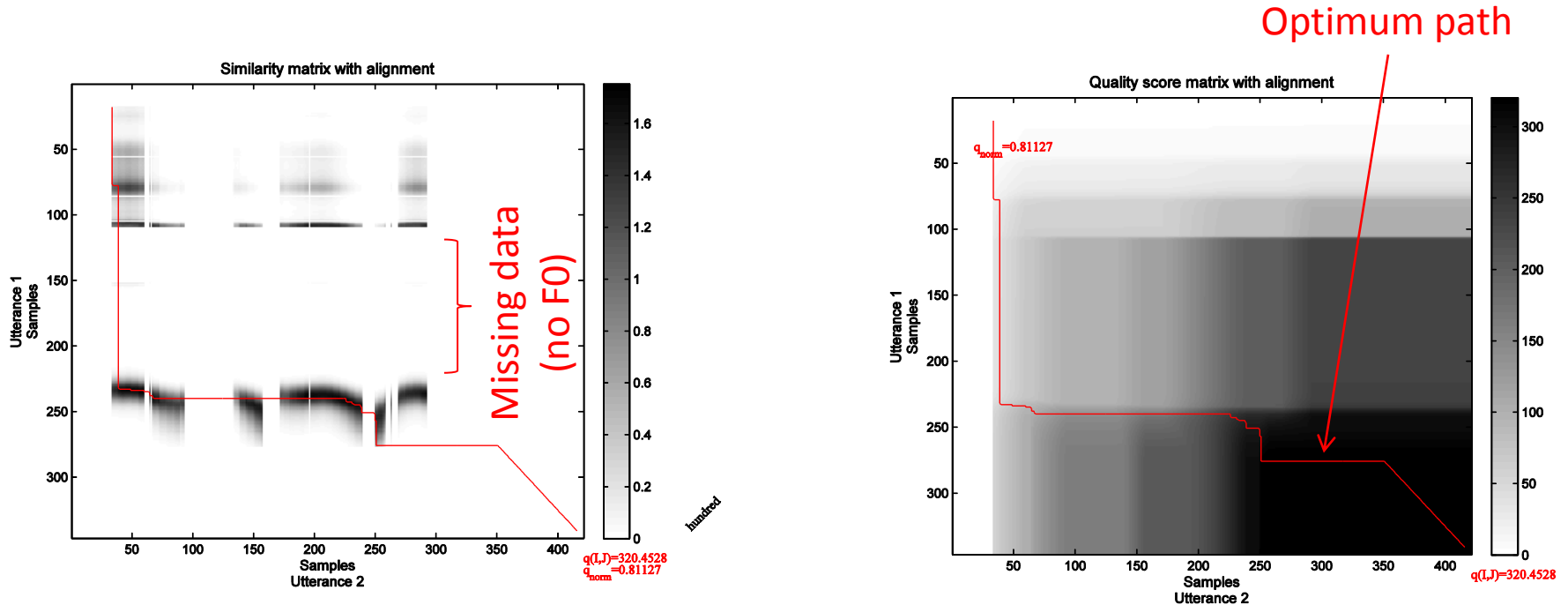


$$\sigma = 0.3$$

[Cooke, 1993]



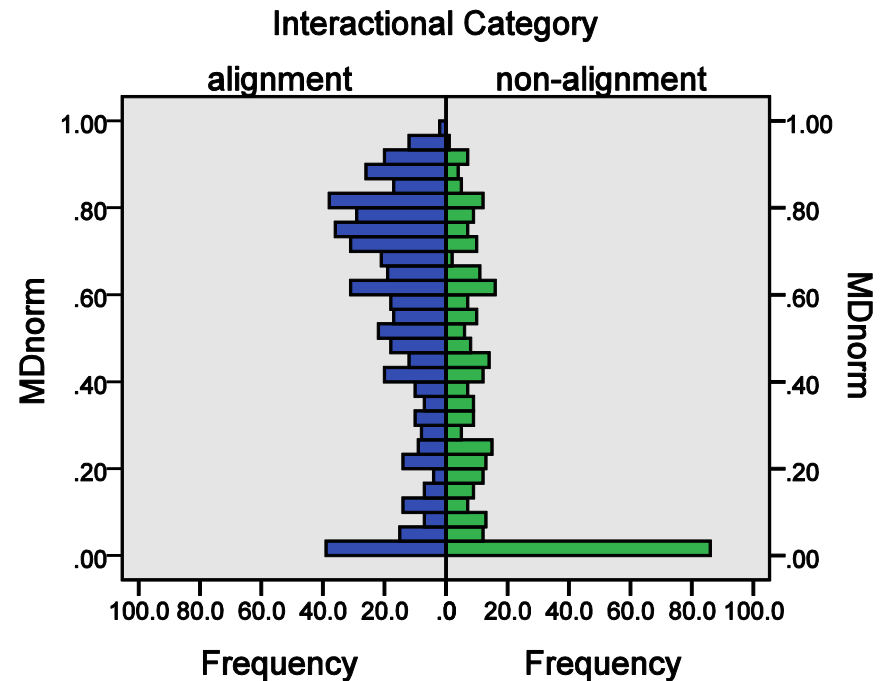
# Accumulative quality score



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

# Results ( $MD_{norm}$ )

- 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

# Outline

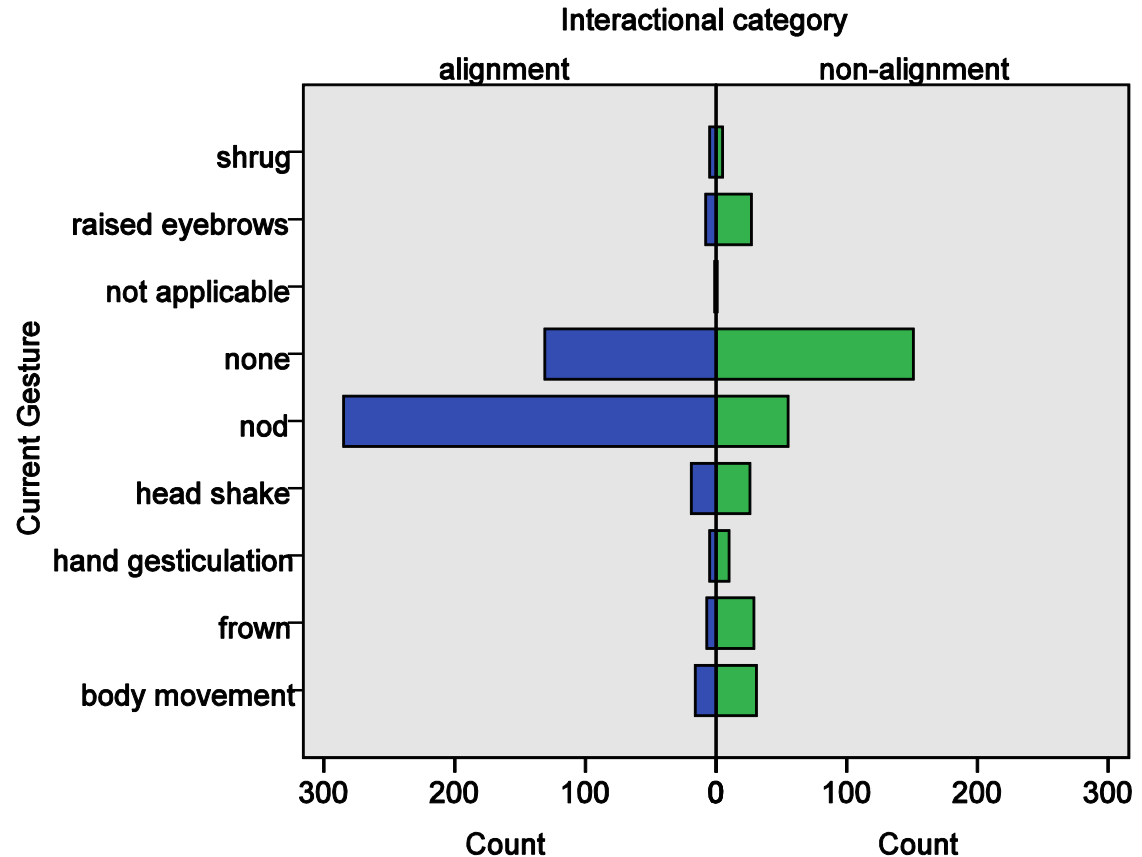
- RQ2a:
  - Are alignments and non-alignments performed with **specific gestures**?
- RQ2b:
  - Are alignments performed with **gestural matches** and are non-alignments performed with **gestural non-matches**?

# H1: specific gestures correlate with interactional category

- Significant interaction between the interactional category and the current gesture:

–  $\chi^2(7, N=810) = 1.686E2, p < 0.001$

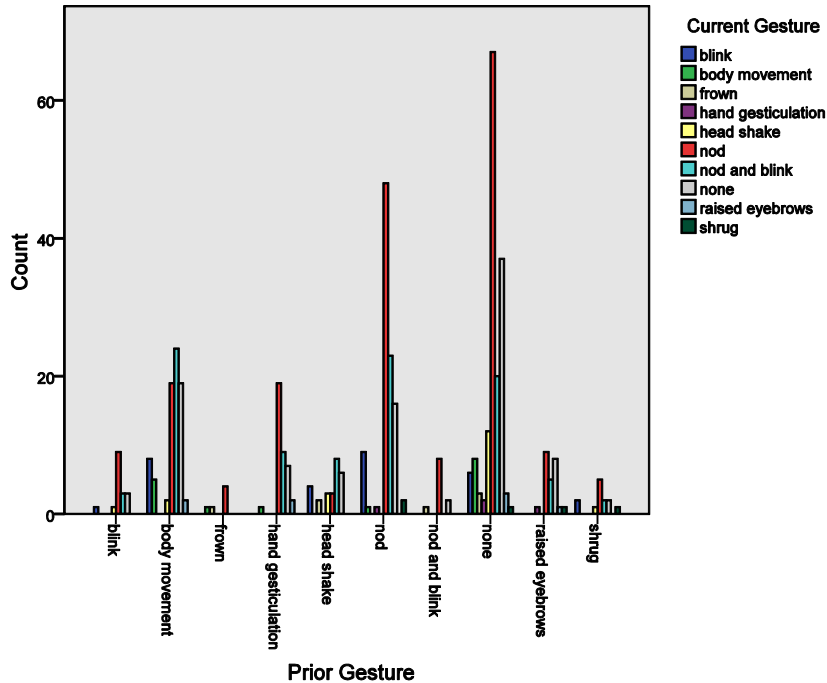
– Effect size:  $\phi = 0.46$



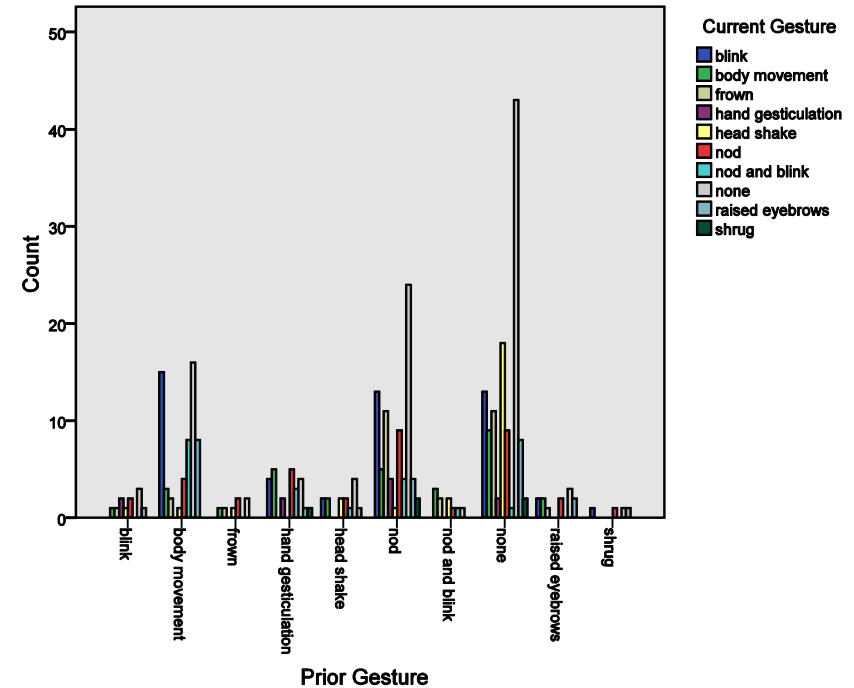
Alignments are more likely to be accompanied by nods

## H2: gestural matching correlates with interactional category

Alignments



Non-alignments



- 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

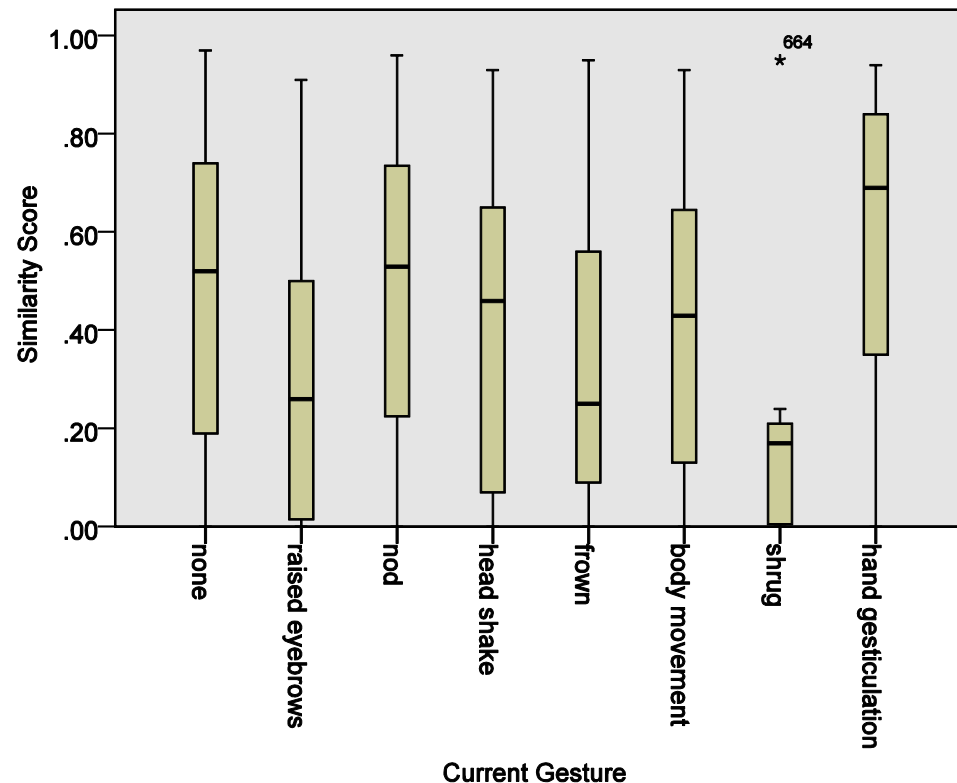
# Outline

- Do gestures correlate with prosodic similarity?
- Combination into acoustic-gestural model
  - Comparison of gesture types and prosodic similarity

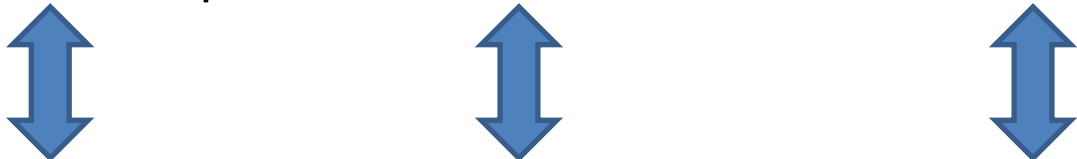
# Prosodic-gestural model

- 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 diagram consists of three blue double-headed vertical arrows. The first arrow is positioned between the first two levels of the analogy, the second between the second and third, and the third between the third and fourth. These arrows indicate a bidirectional relationship or interaction between the levels.

# 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 [Szczepek Reed, 2012b]
- The role of prosodic matching on empathy [Aziz-Zadeh et al., 2010]
- The origin of prosody and gesture



The  
University  
Of  
Sheffield.



# Thank you

[j.gorisch@sheffield.ac.uk](mailto:j.gorisch@sheffield.ac.uk)



# References (1)

- Aziz-Zadeh, L., Sheng, T., & Gheyntanchi, A. (2010). Common Premotor Regions for the Perception and Production of Prosody and Correlations with Empathy and Prosodic Ability. *PLoS ONE*, 5(1), 1-8. doi:10.1371/journal.pone.0008759
- Barth-Weingarten, D. (2011). Double sayings of German JA—more observations on their phonetic form and alignment function. *Research on Language and Social Interaction*, 44 (2), 157-185.
- Cooke, M. P. (1991). Modelling auditory processing and organisation. PhD Thesis. Published by Cambridge University Press, 1993.
- Couper-Kuhlen, E.,(1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen, & M. Selting, *Prosody in conversation: interactional studies* (pp. 366-405). CUP.
- Couper-Kuhlen, E. (2001). Interactional prosody: High onsets in reasons-for-the-call turns. *Language in Society*, 30, 29-53.
- Gardner, R. (2001). *When listeners talk: Response tokens and listener stance* (Vol. 92). (A. H. Jucker, Ed.) Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gorisch, J., Wells, B. & Brown, G. J. (in press). Pitch contour matching and interactional alignment across turns: an acoustic investigation. *Language and Speech*.
- Goodwin, M. H. (1980). Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry*, 50, 303-317.
- Geluykens, R. (1987). Intonation and speech act type: An experimental approach to rising intonation in quecleratives. *Journal of Pragmatics*, 11 (4), 483-494.
- Heath, C. (1992). Gesture's discreet tasks: Multiple relevancies in visual conduct and in the contextualization of language. In P. Auer, & A. di Luzio, *The contextualization of language* (pp. 101-127). Amsterdam: John Benjamins.
- Hermes, D. J. (1998b). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41(1), 73-82.
- Kaimaki, M. (2011). Sequentially determined function of pitch contours: the case of English news receipts. *York Papers in Linguistics* 2, 11, 50-73.
- Kelly, J., & Local, J. (1989). *Doing Phonology: observing, recording, interpreting*. Manchester: Manchester University Press.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Lerner, G. H. (2002). Turn-sharing: the choral co-production of talk-in-interaction. In C. Ford, B. Fox, & S. Thompson, *The Language of Turn and Sequence* (pp. 225-256). Oxford: Oxford University Press.
- Levelt, W. J. (1989). *Speaking*. Cambridge, MA.: The MIT Press

# References (2)

- Local, J., & Wootton, T. (1995). Interactional and phonetic aspects of immediate echolalia in autism: a case study. *Clinical Linguistics & Phonetics*, 9(2), 155-194.
- Müller, F. E. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of reciprocity. In E. Couper-Kuhlen, & M. Selting, *Prosody in conversation: Interactional studies* (pp. 131-176). Cambridge: Cambridge University Press.
- Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38, 1752-1775.
- Ogden, R. (2010). Prosodic constructions in making complaints. In D. Barth-Weingarten, E. Reber, & M. Selting, *Prosody in Interaction* (pp. 81-103). Amsterdam: John Benjamins.
- Rilliard, A., Allauzen, A., & de Mareüil, P. B. (2011). Using dynamic time warping to compute prosodic similarity measures. *Interspeech*. Florence, Italy.
- Schegloff, E. A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50(2), 101-114.
- Selting, M. (2010). Affectivity in conversational storytelling: An analysis of display of anger or indignation in complaint stories. *Pragmatics*, 20(2), 229-277.
- Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41 (1), 31-57.
- Szczepek Reed, B. (2004). Turn-final intonation in English. In E. Couper-Kuhlen, & C. E. Ford, *Sound Patterns in Interaction* (pp. 97-118). Amsterdam: Benjamins
- Szczepek Reed (2006). *Prosodic orientation in English conversation*. Houndmills, Hampshire, UK: Palgrave MacMillan.
- Szczepek Reed, B. (2012a). Beyond the Particular: Prosody and the Coordination of Action. *Journal of Language and Speech*, 55(1), 13-34.
- Szczepek Reed, B. (2012b). Prosody in conversation: Implications for teaching English pronunciation. In J. Romero-Trillo, *Pragmatics, prosody and English language teaching* (pp. 147-168). London: Springer.
- Walker, G. (2004). *The phonetic design of turn endings, beginnings, and continuations in conversation (PhD thesis)*. York: University of York.
- Wells, B. (2010). Tonal repetition and tonal contrast in English carer-child interaction. In D. Barth-Weingarten, E. Reber, & M. Selting, *Prosody in interaction* (pp. 243-262). Amsterdam: John Benjamins.
- Whitehead, K. (2011). Some uses of head nods in 'third position' in talk-in-interaction. *Gesture*, 11(2), 103-122.

end