INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

(Dis)agreement and preference structure are reflected in matching along distinct acoustic-prosodic features

Anneliese Kelterer[1,2], *Margaret Zellers*[3], *Barbara Schuppler*[1]

[1]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria
[2]Institute of Linguistics, University of Graz, Austria
[3]ISFAS, Christian-Albrechts-Universität zu Kiel, Germany

anneliese.kelterer@edu.uni-graz.at, mzellers@isfas.uni-kiel.de, b.schuppler@tugraz.at

Abstract

This paper presents an investigation of acoustic-prosodic alignment in conversational speech and its relationship to functional inter-speaker alignment. While most previous research studied global alignment over whole conversations between strangers, the focus of this paper is on alignment between friends, partners and colleagues as a more local phenomenon related to affiliation and preference structure. Based on 359 turn-pairs from assessment sequences, we analyzed three prosodic matching features between adjacent turns in logistic and linear regression models. We found that disagreements tend to be produced with less F0 span matching than agreements and with less F0 median matching in some parts of the conversation. Preferred responses were more likely to be marked by higher F0 median matching than dispreferred responses. These results indicate that different aspects of functional inter-speaker alignment are reflected in matching along distinct acoustic-prosodic features.

Index Terms: acoustic-prosodic alignment, conversational speech, agreement & disagreement, affiliation, preference structure

1. Introduction

Alignment is a phenomenon in which interlocutors' linguistic productions become closer to each other (also called entrainment, accommodation, mirroring, adaptation; cf., [1, 2]). Prosodic alignment has been found to be related to a variety of social and interactional characteristics, such as likability [3, 4, 5], friendliness [6], attractiveness [5], rapport [7], task success [8] and conversational quality [9, 10]. While most studies have looked at global alignment, that is, how speakers align their prosody to their interlocutor's over the course of a whole conversation, various studies have found that alignment is a dynamic phenomenon [11, 12, 3]. This means that speakers align and disalign at various points in a conversation, related to various conversational functions, such as engagement, conversation flow and symmetry [3]. In line with these findings and with Conversation Analytic (CA) approaches [13, 14], we are not looking at how speakers align to strangers over time (e.g., [15]), but at whether interlocutors who know each other well align their prosody more to each other depending on the particular sequential context. We expect that any initial adaptation between these speakers has already been completed and that they align more or less depending on whether they are more or less 'in sync' at that moment in the conversation.

Thus, we investigate prosodic alignment between adjacent turns in assessment sequences in Austrian German spontaneous conversations between friends, partners and colleagues. Assessment sequences are adjacency pairs in which the first assessment about something opens up the floor for another speaker to utter an assessment about the same thing, thereby affiliating or disaffiliating with them [16]. Automatic entrainment may of course still play a role, but we assume that it would already have been completed for the speakers investigated here, because they know each other very well.

Since different studies have found alignment along different phonetic parameters, Ostrand & Chodroff [17] have suggested that entrainment is not a uniform phenomenon in which all features entrain, and that discrepancies in findings for different acoustic features might be related to different functions. Therefore, we look at three acoustic features that express different prosodic percepts: F0 span (is both speakers' intonation equally variable or monotone?), F0 median (do both speakers talk in the same register?) and articulation rate (do interlocutors match their speed of talking?). The two F0 features may correlate, but need not do so; for instance, if one speaker has a flat intonation in the middle of their range and the other a very variable intonation in the middle of their range, F0 median matching is high but F0 span matching is low. We expect prosodic alignment to be a phenomenon that results from speakers normalizing for their interlocutor's F0 range. In other words, we expect that alignment along F0 span and median is a phenomenon related to speakers' individual F0 ranges rather than to absolute F0 (cf., absolute pitch matching in mimicry [18]).

In this paper, we investigate two aspects of local functional alignment. Extrapolating from findings for rapport (e.g., [7]), the first is *affiliative stance* (i.e., whether speakers agree or disagree with their interlocutor; called affiliation in the remainder of the paper). Previous task-based studies do not suggest a strong relationship between prosodic alignment and affiliation [11, 12]. In a study of spontaneous conversation, however, Szczepek Reed [14] found a considerable number of prosodically aligned agreements compared to a very small number of disagreements, even though she argues that affiliation is not the main reason why interlocutors mirror each other in spontaneous conversations (see below). Thus, we hypothesize that interlocutors match their prosody more in agreeing than in disagreeing turns (Hypothesis 1).

The second aspect of local functional alignment we investigate is *preference structure*. In adjacency pairs (such as assessment sequences or invitations) in which interlocutors have a choice between different second pair parts (e.g., agreement vs. disagreement, acceptance vs. rejection), one of these responses is *structurally* preferred over the other (i.e., regardless of personal preference of the interlocutors) [19]. The preferred option has been described as the one that is invited over its alternative [20], ensures the progression of the action in progress [19], expresses sociability, support and solidarity [21] or the one that is relevantly absent if it is not uttered [22]. Interlocutors orient to this preference structure by, for instance, delaying

10.21437/Interspeech.2023-1538

dispreferred responses [20, 23] and producing them with downgraded prosody [16]. In this study, we look only at action type preference as defined above, not at 'preferred' (i.e., immediate, phonetically upgraded) vs. 'dispreferred' (i.e., delayed, phonetically downgraded) turn formats (cf., [24]).

In assessment sequences, agreement is preferred in friendly conversations [20, 25, 19], and disagreement is preferred in argumentation [26], after self-deprecations [20], and after accusations and attributions made of others [27] (e.g., "you said…", "you wanted…"). We are not aware of any previous studies explicitly relating prosodic matching to action type preference, but a few studies have investigated phenomena that are relevant to this question, though their findings are mixed. Ogden [16] found different kinds of non-matching depending on preference structure; prosodic upgrading in preferred and downgrading in dispreferred responses. However, he did not compare whether speakers' productions were closer to their interlocutor's productions in preferred or dispreferred responses. Weidman et al. [9] found that romantic partners aligned their prosody less during conflict. Szczepek Reed [14] argued that prosodic matching is used to "establish coherent trajectories of action" (p. 132) (see definition of preference by [19] above) rather than to affiliate (cf. also [13]). In an argument, for instance, a disagreement follows the discourse trajectory and would be expected to display more matching. In the same context, an agreeing turn may be the beginning of a new sequence (e.g., reconciliation with subsequent topic change) and would therefore not be expected to display prosodic matching. Thus, we hypothesize that interlocutors match their prosody more in preferred responses than in dispreferred responses (Hypothesis 2).

2. Methods

2.1. Corpus and annotation process

The data come from six dyadic conversations from GRASS [28, 29] between 12 native speakers of Austrian German (2m-m, 4f-m) who were friends, partners or colleagues. The one-hour long conversations were recorded in a recording studio, while no experimenter was present and no topic was given. We identified assessment sequences as a sequence of turn 1 and turn 2 in which both expressed an assessment about the same referent [20] (cf., stance object in [30]), though the referent might also be negotiated over time, for instance, in the course of an argument. A speaker might, for instance, produce a partial agreement with an interlocutor's assessment, conceding a minor point, but continue the disagreement about the main point [26], which modifies the referent. Even though the referent is not the same in these cases, we still included them in our data. All annotations were done by the first author. Based on annotations of 40 assessment sequences annotated by the first and second author and a colleague, annotation criteria were revised and all annotations corrected according to the updated criteria.

Agreements were identified when both interlocutors communicated the same opinion about the referent. Disagreements were identified when the second assessment included an antonym (e.g., *boring - good*; *weird - not bad* when talking about a movie; A: "it was weird, wasn't it?" - B: "no, it wasn't bad at all"), had opposite polarity (e.g., *I like it - I don't*) (cf., [16]), or offered an alternative (e.g. talking about why the pope stepped down; A: "maybe he wanted to go down in history" - B: "no, he just looks old").

To assess preference structure, turn pairs were labelled according to the context in which they were uttered. Since all conversations in GRASS are friendly at the outset and no topics were given, discussions and other situations in which disagreement is preferred occurred spontaneously in these conversations. It is thus important to have an instrument to identify sequences in which disagreement is the preferred action. We assumed that the default in these conversations is that agreement is preferred (cf., Table 1) and identified three contexts in which disagreement is preferred: a) re-statements of an opposing opinion [26, 16], b) in arguments (when the opposition continues after the first re-statement; cf., [26]), c) after attributions and assumptions about the interlocutor (e.g., "you did…", "you said…", "you wanted…", etc.), including accusations [22, 27]. The preference status of agreement and disagreement in these situations is summarised in Table 1. In addition to these contexts, Pomerantz [20] observed that after compliments and self-deprecating comments, agreement was not preferred, but we did not find these contexts in our data.

Table 1: *Contexts in which agreement (AG) and disagreement (DG) were labelled as preferred or dispreferred.*

	preferred	dispreferred
AG	friendly context (default) N = 143	arguments, re-statements of opposing opinion, after other-attributions N = 40
DG	arguments, re-statements of opposing opinion, after other-attributions N = 85	friendly context (default) N = 89

2.2. Domain of feature extraction and data set

Prosodic matching is measured within the assessment pair, that is, between assessments in two adjacent turns. In longer assessment sequences, in which, for instance, two disagreements were followed by a concession (i.e., an agreement), prosodic matching between each (dis)agreement and the previous assessment turn was measured. This differs from the method in [14], where all assessments in longer sequences were counted as one instance.

To allow the prosodic comparison of two assessments, they should be as comparable as possible. One issue for assessing prosodic matching is when the two assessments are of very different lengths, since in very short turns, the same degree of variation is not possible as in larger turns. To make the two assessments as comparable as possible, one- and two-word assessments were excluded (e.g. "arg" *terrible*, "voll cool" *really cool*, "ja genau" *yeah exactly*). In very long turns, we limited the domain for prosodic feature extraction to the minimum number of Turn Construction Units that still expressed the stance at the end of assessment 1 (given that that the referent of the assessment had already been expressed, i.e., that it did not have to be inferred) and at the beginning of assessment 2 (but more than just assessment terms like "ja", "voll", or "nein").

In addition to one- and two-word assessments, we excluded tokens with laughed speech, where the preference context was unclear, and where the timing between the first and the second assessment was ambiguous (e.g., in simultaneous speech, or when the first speaker uttered several assessments and it was not clear which of those the second speaker (dis)agreed with, often in (partial) overlap). This resulted in a data set of 184 agreement (144 preferred, 40 dispreferred) and 175 disagreement (85 preferred, 90 dispreferred) assessment pairs.

2.3. Matching features

We calculated matching along three parameters: F0 span (*match(F0$_{span}$)*), F0 median (*match(F0$_{med}$)*) and articulation rate (*match(AR)*). F0 was manually corrected and smoothed with mausmooth [31] and then converted to semitones based on the median of each speaker over the whole conversation. F0 span and F0 median were then calculated within the domain of each assessment in the assessment pair (cf., Section 2.2). Phone segmentations consisted of manually corrected forced alignments from Kaldi [32]. AR was calculated as the number of phones per second in the same domain, not considering any pauses.

For all three features, matching was calculated as $(-\Delta(turn1, turn2))$, that is, the difference between this feature in turn 1 and turn 2. A lower value indicates less matching and a value closer to zero indicates more matching. Pearson's Correlation Coefficients were low for the three matching features (*match(F0$_{span}$)* x *match(F0$_{med}$)*: r(355) = .055, p = .303; *match(F0$_{span}$)* x *match(AR)*: r(355) = .064, p = .230; *match(F0$_{med}$)* x *match(AR)*: r(355) = -.001, p = .988).

2.4. Statistical methods

To gain insights about how the acoustic matching features contribute to the expression of affiliation (categorical variable *Affiliation* with the values AG and DG; cf., first vs. second row in Table 1) and to the expression of *Preference* (with the values preferred and dispreferred; cf., left vs. right column in Table 1), we built two separate mixed effects logistic regression models, one predicting *Affiliation* and one predicting *Preference*, using the lme4 package in R [33]. The maximum combination of independent variables were the matching features *match(F0$_{span}$)*, *match(F0$_{med}$)* and *match(AR)*, as well as their two-ways interactions, and *Speaker* as a random effect (the speaker producing the second assessment in each pair; n = 12).

To analyze how the acoustic matching features *match(F0$_{span}$)*, *match(F0$_{med}$)* and *match(AR)* are related to both *Affiliation* and *Preference* in more detail, we built linear mixed effects regression models with the lme4 package in R. To account for possible long-term effects (e.g., due to tiredness), we calculated the relative position of each assessment in the one-hour long conversations (normalized between 0 and 1; *Position*). The maximum model contained the respective matching feature as the dependent variable, *Affiliation*, *Preference* and *Position* as independent variables, as well as their two-way interactions, and *Speaker* as a random effect .

To obtain the best models, we performed best subsets regression and selected the model with the lowest AIC [34]. Only these models are presented in Section 3.

3. Results

3.1. Which matching features predict affiliation and preference structure?

The final model predicting *Affiliation* (cf., Hypothesis 1) included only *match(F0$_{span}$)* and *Speaker* as predictors of *Affiliation* and is given in Table 2. When *match(F0$_{span}$)* is higher, DG is less likely than AG. In other words, AG is more likely when F0 span matches to a higher degree and DG when F0 span shows less matching. Marginal and conditional R^2 indicate that *match(F0$_{span}$)* explains 2% and the random effect explains an additional 10% of variance in this model.

The final model predicting *Preference* (cf., Hypothesis 2) includes only *match(F0$_{med}$)* and is given in Table 3. When

Table 2: *Final model predicting Affiliation (levels: AG vs. DG). Marginal R^2 = 0.02, conditional R^2 = 0.12.*

	Est.	z	p
(Intercept)	-0.53	-2.05	$< .05$
match(F0$_{span}$)	-0.11	-2.73	$< .01$
Speaker		sd = 0.66	

match(F0$_{med}$) is higher, the dispreferred response is less likely than the preferred response. In other words, the preferred response is more likely at higher F0 median matching values and the dispreferred response at lower values. R^2 indicates that *match(F0$_{med}$)* explains 2% of variance in this model.

Table 3: *Final model predicting Preference (levels: preferred vs. dispreferred response). Marginal R^2 = 0.02.*

	Est.	z	p
(Intercept)	-0.99	-5.22	$< .001$
match(F0$_{med}$)	-0.22	-2.8	$< .01$

3.2. What effects do affiliation and preference structure have on prosodic matching features?

The final model for *match(F0$_{span}$)* included the independent variables *Affiliation* and *Position*, but no interaction (cf., Table 4). R^2 indicates that 5% of variance was explained by these variables. DG had significantly lower *match(F0$_{span}$)* than AG and *match(F0$_{span}$)* decreased towards the end of the conversation.

Table 4: *Final model predicting match(F0$_{span}$). Marginal R^2 = 0.05.*

	Est.	t	p
(Intercept)	-2.07	-6.57	$< .0001$
Affiliation (DG)	-0.94	-3.24	$< .01$
Position	-1.28	-2.49	$< .05$

The final model for *match(F0$_{med}$)* included the independent variables *Affiliation*, *Preference*, *Position* and an interaction of *Affiliation* and *Position* (cf., Table 5). R^2 indicates that 4% of variance are explained by this model. *match(F0$_{med}$)* was significantly lower in dispreferred responses. There was no significant difference between AG and DG at the beginning of the conversation and *Position* alone was not significant. The significant interaction between *Affiliation* and *Position*, however, indicates that *match(F0$_{med}$)* was lower for DG than AG at the end of the conversation.

Table 5: *Final model predicting match(F0$_{med}$). Marginal R^2 = 0.04.*

	Est.	t	p
(Intercept)	-1.93	-9.66	$< .0001$
Affiliation (DG)	0.43	1.42	.155
Preference (dispref.)	-0.41	-2.53	$< .05$
Position	0.57	1.62	.106
Affiliation (DG) x *Position*	-1.27	-2.42	$< .05$

In the final model for articulation rate matching, *match(AR)* was predicted only by *Position*, indicating a slight rise in articulation rate matching, but *Position* only attained a threshold of smaller than .1 (Est. = 0.8977, t = 1.882, $p < .061$). R^2 indicates that this model only explains 1% of variance. Thus, we conclude that there is no relationship between articulation rate matching and either *Affiliation* or *Preference* in our data.

4. Discussion

In this study, we assessed the relationship between prosodic alignment along three parameters (F0 span, F0 median and articulation rate) with two kinds of functional inter-speaker alignment. Despite the immense variation in the spontaneous conversational speech investigated here (reflected in relatively low R^2 values of the fitted models), we did find patterns of prosodic alignment.

The logistic regression models predicting *Affiliation* and *Preference* showed that the two annotated alignment dimensions were predicted by different matching features and no interaction of these features was included in the final models. As expected, we found more matching in agreement than in disagreement, but only for the feature F0 span (Hypothesis 1), and more matching in preferred than in dispreferred second pair parts, but only for the feature F0 median (Hypothesis 2). Articulation rate matching was not relevant in the prediction of either *Affiliation* or *Preference*.

The linear model predicting *match($F0_{span}$)* mirrors the role of *match($F0_{span}$)* in the logistic models in that it showed that F0 span matching was higher in AG than in DG (Hypothesis 1), but that it was not related to preference structure. The linear model predicting *match($F0_{med}$)* also confirms the relationship between *match($F0_{med}$)* and preference shown in the logistic model and was higher for preferred than for dispreferred responses (Hypothesis 2). However, this linear model gives us additional insights into a relationship between F0 median matching and *Affiliation* (cf. Hypothesis 1) related to time. We found no F0 median matching difference between AG and DG at the beginning of conversations, but significantly less median matching in DG towards the end. One reason for this relationship with time could be that interlocutors cared more about mitigating their disagreements at the beginning of conversations. However, all interlocutors knew each other very well, so an adjustment period similar to conversations between strangers is unlikely. Another possible reason is that the relationship to time is an artefact and the degree of F0 median matching in agreement vs. disagreement is related to the specific contexts and topics speakers (dis)agree about which just happen to occur more towards the beginning vs. the end of the conversations in our data. A more detailed investigation of contexts (e.g., arguments vs. attributions and assumptions about the interlocutor, cf., [27]) and particular topics (e.g., epistemic discussions vs. discussions about personal attitudes) could shed more light on this issue.

The linear model predicting articulation rate confirmed the results of the logistic models, finding no significant relationship between articulation rate and either *Affiliation* or *Preference* in our data. This corresponds to other studies that did not find a relationship between articulation rate alignment and any functional aspects of alignment [3, 12].

Concerning affiliation, our findings are in line with Szczepek Reed's [14] findings of a higher number of matching agreements than disagreements. One reason why Bonin et al. [12] and Vaughan [11] did not find a consistent relationship of affiliation to prosodic alignment while we did might be due to the different data collection settings [4, 35]. Bonin et al. and Vaughan studied task-based dialogues in which interlocutors could not see each other, while we investigated spontaneous conversations (i.e., without any tasks given) between friends, partners and colleagues who were sitting in the same room.

Our findings of more F0 median matching in preferred responses indicate that speakers match more in agreements in the default friendly context as well as in disagreements in arguments (left column in Table 1). This stands in contrast to Weidman et al.'s [9] findings of overall less matching during conflict, though we did not look at assessments in arguments vs. other situations (diagonal comparison in Table 1), so our study is not entirely comparable. Our results of more matching in preferred responses are, however, in line with Szczepek Reed's [14] investigation, even though she did not explicitly study preference structure. She found many matching instances in agreements, but also in disagreements that occur in longer disagreement sequences. Her examples also include re-statements and one response to an attribution about the interlocutor, all of which would be characterized as preferred disagreements in our annotation scheme (cf., Section 2.1). Interestingly, in the one example of such a longer sequence she gives, interlocutors also mostly matched in terms of pitch register.

We found more matching, that is, a tendency for the absence of either up- or down-grading for F0 span and F0 median in agreements and for F0 median preferred responses. Whether disagreements and dispreferred responses (in which we found less matching) are characterised by up- or down-grading [16] is a question for future work.

In light of the considerable speaker and dyad variation found in the literature (e.g., [36, 37, 38, 35]), it is notable that speaker was significant only in one of the models examined here. Speaker variation will be investigated further in the future, though it may be that relevant variation is on the level of the dyad [39], or related to intra-speaker variation. CA, for instance, describes behaviour in terms of an inventory of strategies, and speakers can choose which of these strategies to use to achieve a communicative goal. For instance, engagement could play a role [3], revealing different matching behaviour in heated vs. calm discussions or enthusiastic vs. token agreement.

To summarize, our results confirm that entrainment is not a uniform phenomenon and that different functions are related to matching along different acoustic-prosodic parameters [17]. When expressing a disaffiliative stance (i.e., a disagreement), speakers tend to match their interlocutors less than when affiliating (i.e., agreeing) in terms of how 'lively' their intonation is and in some parts of the conversations also in their register. When producing a dispreferred response in assessment sequences, speakers tend to match the register of their interlocutors less than in preferred responses.

5. Conclusion

In this study, we investigated the relationship between functional and formal alignment in assessment sequences in casual Austrian German conversations between partners, friends and colleagues. We found overall more F0 span matching and more F0 median matching in some parts of the conversation in agreement than in disagreement, and more F0 median matching in preferred than in dispreferred second pair parts, but no relationship of either affiliation or preference structure with articulation rate matching. Our results show that different matching features may be associated with different functional aspects of interpersonal alignment.This provides further evidence for the importance of a) investigating alignment along different acoustic-prosodic parameters, and b) of looking at affiliation and preference structure separately.

6. Acknowledgements

The work by A. Kelterer was funded by grant P-32700-NB from FWF (Austrian Science Fund).

7. References

[1] C. J. Wynn and S. A. Borrie, "Classifying conversational entrainment of speech behavior: An expanded framework and review," *Journal of Phonetics*, no. 94, pp. 1–11, 2022.

[2] J. S. Pardo, E. Pellegrino, V. Dellwo, and B. Möbius, "Special Issue: Vocal accommodation in speech communication," *Journal of Phonetics*, no. 95, pp. 1–9, 2022.

[3] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell, "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction," *Speech Communication*, no. 58, pp. 11–34, 2014.

[4] K. Schweitzer, M. Walsh, and A. Schweitzer, "To see or not to see: Interlocutor visibility and likeability influence convergence in intonation," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden, Aug. 2017, pp. 919–923.

[5] J. Michalsky and H. Schoormann, "Pitch convergence as an effect of perceived attractiveness and likability," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden, Aug. 2017, pp. 2253–2256.

[6] M. Zellers and A. Schweitzer, "An investigation of pitch matching across adjacent turns in a corpus of spontaneous German," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden, Aug. 2017, pp. 2336–2340.

[7] N. Lubold and H. Pon-Barry, "Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues," in *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics*, 2014, pp. 5–12.

[8] Z. Rahimi, A. Kumar, D. Litman, S. Paletz, and M. Yu, "Entrainment in Multi-Party Spoken Dialogues at Multiple Linguistic Levels," in *Proc. INTERSPEECH 2017*, Stockholm, Sweden, Aug. 2017, pp. 1696–1700.

[9] S. Weidman, M. Breen, and K. C. Haydon, "Prosodic Speech Entrainment in Romantic Relationships," in *Proceedings of Speech Prosody 2016*, Boston, USA, May 2016, pp. 508–512.

[10] J. Michalsky, H. Schoormann, and O. Niebuhr, "Conversational quality is affected by and reflected in prosodic entrainment," in *Proceedings of Speech Prosody 2018*, Poznań, Poland, June 2018, pp. 389–392.

[11] B. Vaughan, "Prosodic Synchrony in Co-operative Task-based Dialogues: A Measure of Agreement and Disagreement," in *Proc. INTERSPEECH 2011*, Florence, Italy, Aug. 2011, pp. 1865–1868.

[12] F. Bonin, C. De Looze, S. Ghosh, E. Gilmartin, C. Vogel, A. Polychroniou, H. Salamin, A. Vinciarelli, and N. Campbell, "Investigating fine temporal dynamics of prosodic and lexical accommodation," in *Proc. INTERSPEECH 2013*, Lyon, France, Aug. 2013, pp. 539–543.

[13] J. Gorisch, B. Wells, and G. J. Brown, "Pitch Contour Matching and Interactional Alignment across Turns: An Acoustic Investigation," *Language and Speech*, vol. 1, no. 55, pp. 57–76, 2012.

[14] B. Szczepek Reed, "Reconceptualizing mirroring: Sound imitation and rapport in naturally occurring interaction," *Journal of Pragmatics*, no. 167, pp. 131–151, 2020.

[15] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proc. INTERSPEECH 2011*, Florence, Italy, Aug. 2011, pp. 3081–3084.

[16] R. Ogden, "Phonetics and social action in agreements and disagreements," *Journal of Pragmatics*, no. 38, pp. 1752–1775, 2006.

[17] R. Ostrand and E. Chodroff, "It's alignment all the way down but not all the way up: Speakers align on some features but not others within a dialogue," *Journal of Phonetics*, no. 88, pp. 1–17, 2021.

[18] E. Couper-Kuhlen, "The prosody of repetition: on quoting and mimicry," in *Prosody in conversation*, E. Couper-Kuhlen and M. Selting, Eds. Cambridge University Press, 2012, pp. 366–405.

[19] E. A. Schegloff, "The organization of preference/dispreference," in *Sequence Organization in Interaction: A Primer in Conversation Analysis*, E. A. Schegloff, Ed. Cambridge University Press, 2007, pp. 58–96.

[20] A. Pomerantz, "Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes," in *Structures of Social Action*, J. M. Atkinson and J. Heritage, Eds. Cambridge University Press, 1984, pp. 57–101.

[21] A. Pomerantz and J. Heritage, "Preference," in *The Handbook of Conversation Analysis*, J. Sidnell and T. Stivers, Eds. Wiley-Blackwell, 2012, pp. 210–228.

[22] J. Bilmes, "The Concept of Preference in Conversation Analysis," *Language and Society*, vol. 2, no. 17, pp. 161–181, 1988.

[23] K. H. Kendrick and F. Torreira, "The Timing and Construction of Preference: A Quantitative Study," *Discourse Processes*, no. 52, pp. 255–289, 2015.

[24] D. Pillet-Shore, "Preference Organization," in *Oxford Research Encyclopedia of Communication*. Oxford University Press, 2017.

[25] H. Sacks, "On the Preferences for Agreement and Contiguity in Sequences in Conversation," in *Talk and Social Organisation*, G. Button and J. R. Lee, Eds. Clevedon: Multilingual Matters, 1987, pp. 54–69.

[26] H. Kotthoff, "Disagreement and Concession in Disputes: On the Context Sensitivity of Preference Structures," *Language in Society*, no. 22, pp. 193–216, 1993.

[27] J. Bilmes, "Preference and the conversation analytic endeavor," *Journal of Pragmatics*, no. 64, pp. 52–71, 2014.

[28] B. Schuppler, M. Hagmüller, J. A. Morales-Cordovilla, and H. Pessentheiner, "GRASS: the Graz corpus of Read And Spontaneous Speech," in *Proc. LREC*, 2014, pp. 1465–1470.

[29] B. Schuppler, M. Hagmüller, and A. Zahrer, "A corpus of read and conversational Austrian German," *Speech Communication*, no. 94, pp. 62–74, 2017.

[30] J. W. Du Bois, "The stance triangle," in *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, R. Englebretson, Ed. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007, pp. 139–182.

[31] F. Cangemi, "mausmooth," 2015. [Online]. Available: http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

[32] D. Povey, A. Goschal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The Kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and unterstanding*, 2011.

[33] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 1, no. 67, pp. 1–48, 2015.

[34] J. G. Liao, J. E. Cavanaugh, and T. L. McMurry, "Extending AIC to best subset regression," *Computational Statistics*, no. 33, pp. 787—-806, 2018.

[35] J. S. Pardo, A. Urmanche, S. Wilman, J. Wiener, N. Mason, K. Francis, and M. Ward, "A comparison of phonetic convergence in conversational interaction and speech shadowing," *Journal of Phonetics*, no. 69, pp. 1–11, 2018.

[36] R. Levitan, Š. Beňuš, A. Gravano, and J. Hirschberg, "Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison," in *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Prague, Czech Republic, Sep. 2015, pp. 325–334.

[37] A. Menshikova, D. Kocharov, and T. Kachkovskaia, "Phonetic Entrainment in Cooperative Dialogues: A Case of Russian," in *Proc. INTERSPEECH 2020*, Shanghai, China, Oct. 2020, pp. 4148–4152.

[38] P. Šturm, R. Skarnitzl, and T. Nechanský, "Prosodic Accommodation in Face-to-face and Telephone Dialogues," in *Proc. INTERSPEECH 2021*, Brno, Czechia, Aug.-Sep. 2021, pp. 1444–1448.

[39] A. Paxton and R. Dale, "Argument disrupts interpersonal synchrony," *The Quarterly Journal of Experimental Psychology*, vol. 11, no. 66, pp. 2092–2102, 2013.