*INTERSPEECH 2017*
August 20–24, 2017, Stockholm, Sweden

# An investigation of pitch matching across adjacent turns in a corpus of spontaneous German

*Margaret Zellers*[1], *Antje Schweitzer*[2]

[1]Institut für Linguistik: Anglistik, Universität Stuttgart, Germany
[2]Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany

margaret.zellers@ifla.uni-stuttgart.de, antje.schweitzer@ims.uni-stuttgart.de

## Abstract

Speakers in conversations may adapt their turn pitch relative to that of preceding turns to signal alignment with their interlocutor. However, the reference frame for pitch matching across turns is still unclear. Researchers studying pitch in the context of conversation have argued for an initializing approach, in which turn pitch must be judged relative to pitch in preceding turns. However, perceptual studies have indicated that listeners are able to reliably identify the location of pitch values within an individual speaker's range; that is, even without conversational context, they are able to normalize to speakers. This would imply that speakers might match normalized pitch instead of absolute pitch. Using a combined quantitative-qualitative approach, we investigate the relationship between pitch in adjacent turns in spontaneous German conversation. We use two different methods of evaluating pitch in adjacent turns, reflecting normalizing and initializing approaches respectively. We find that the results are well correlated with conversational participants' evaluation of the conversation. Furthermore, evaluating locations with matched or mismatched pitch can help distinguish between blind and face-to-face conversational situations, as well as identifying locations where specific discourse strategies (such as tag questions) have been deployed.

**Index Terms**: interaction, pitch, turn-taking, German

## 1. Introduction

Pitch has been reported by many researchers to provide contextualization information which allows listeners to infer meaning of utterances beyond their lexical content and make "situated interpretations" [1]. The situation in which these interpretations occur is the surrounding turns; the position and the form of a conversational turn are both relevant in understanding what a speaker is doing [2, 3]. Among many other functions, pitch can contribute to communicating discourse-internal factors such as solidarity among speakers [4] and agreement or disagreement [5]. Furthermore, speakers may accommodate or converge prosodically to one another during some phases of discourse, although the exact mechanisms underlying such accommodation are still unclear [6, 7, 8, 9].

Listeners adapt quickly to their interlocutors' specific voice features. For example, they are able to reliably identify the location of fundamental frequency (F0) values relative to an individual speaker's range [10]. While this may be due in part to listeners' expectations regarding average F0 for different speaker sexes [11], voice quality also contributes to the perception of pitch height, and manipulating voice quality can manipulate listeners' judgments about the location of F0 values within a speakers pitch range [12]. Thus it appears that even without conversational context, listeners have some ability to rank pitch as high or low; that is, they *normalize* for a speaker [13]. However, researchers studying pitch specifically in the context of conversation often argue for an approach in which turn pitch must be judged relative to what has come before in interaction; [13] calls this an *initializing* approach. Recent research into the production of prosody in adjacent turns has provided support for both normalizing and initializing views. [14] find that pitch contours in backchannels closely match the pitch contour at the end of the turn they follow, which is difficult to account for using normalization. [15], on the other hand, report that the function of turns with question structure can be identified by whether they have high or low initial pitch (i.e. not near the speaker's median pitch, and thus considered to mismatch with the context), without reference to the preceding turn's pitch.

While [14] and [15] report on pitch features at turn edges, other studies have considered more internal features of turns. [16] considered global prosodic features of adjacent turns in American English, specifically those containing contrastive structures. They found that whether or not global prosodic features between turns matched was related to the degree of "problematicness" that the interaction involved; more problematic turn pairs were more closely matched prosodically than turn pairs which did not appear to entail any interactional difficulty.

The current study investigates pitch matching using pairs of full turns (see criteria in Section 2.2) in spontaneous conversation, and takes into account the syntactic structure and discourse features of the turns. We use two methods of interpreting turn pitch, designed to reflect the normalizing and initializing approaches described above. Thus our quantitative analysis is informed by qualitative observation of the interactive situation.

## 2. Data and methodology

### 2.1. GECO corpus

The data used for this study come from the GECO corpus [17, 18], which is comprised of 46 spontaneous conversations of approximately 25 minutes each, between previously unacquainted female German speakers. After each dialog, participants filled in questionnaires about their assessment of the dialog and their impression of the conversation partner. Questions referring to the assessment of the dialog included aspects such as whether the dialog was friendly, or stressful, or whether participants experienced any communication problems. Questions referring to the impression of the conversation partner asked whether the partner was, for example, likeable, friendly, or intelligent, on a scale of -2 (not at all) to 2 (very). 22 of the dialogs were recorded in such a way that speakers could not see each other ("monomodal" condition), while the remaining



http://dx.doi.org/10.21437/Interspeech.2017-811

24 dialogs were recorded with the speakers sitting face-to-face[1] ("multimodal" condition). All speakers were involved in several dialogs. Altogether the corpus consists of approximately 21 hours of conversational speech on subjects of the speakers' choice. All dialogs were manually transcribed and then automatically labeled on the word, syllable, and segment levels using forced alignment.

### 2.2. Automatic identification of turn boundaries

We automatically identified turn boundaries based on the word-label files. We identified utterances as stretches of speech by one speaker that were not interrupted by pauses longer than 0.3 s. Turns were then defined as utterances with a minimum length of 1.5 s. We automatically identified turn transitions where the next turn did not overlap with the current turn by more than 0.5 s, and where the next turn was not overlapped by another utterance (except for utterances shorter than 0.5 s, which we treated as backchannels). We discarded turns with more overlap in order to avoid cases where one speaker fails to yield the floor, or where both speakers attempt to take the floor simultaneously. In such cases, we expect that there is no motivation for the speakers to match their turns prosodically; on the contrary, it is likely that speakers will use pitch in order to hold or take the floor (cf. [19]). Turns with laughter were discarded because its pitch typically exceeds that of normal conversational speech and thus might have corrupted our pitch analysis.

### 2.3. Classification of turn transition pitch values

This study investigates whether and in what contexts speakers match their register at speaker transitions, i.e. whether speakers who take the floor will match the preceding speaker's register. To this end all turns at turn boundaries were classified according to their register. We distinguished low, mid, and high register, thus yielding 9 classes of turn transitions ranging from low-low, low-mid, low-high, ..., to high-high. We used two different methods to classify the turns. The first method adheres to a normalizing view, assuming that the second speaker will perceive the partner's register normalized to her pitch range and match that in her own register. The second method, on the other hand, suggests that the speaker taking the floor perceives the preceding register relative to her own pitch range, and thus represents an initializing view.

To calculate the pitch of each speaker in each dialog, we modified the procedure suggested by [20]. We first calculated pitch for each dialog and each speaker with Praat [21] using a pitch range of 75 to 600 Hz. Following [20], we then made a second pass using 0.75 times the first quartile as a new lower bound, and 2 times the third quartile as a new upper bound (instead of 1.5 times, as suggested by [20] for read speech). We found this to be more appropriate for the conversational speech in the GECO database, which sometimes exhibits rather high pitch values, often in more emotional stretches. In a third pass, we again adjusted the lower bound to 0.75 times the first quartile, but since we found that the maximum values were already mostly correct, we took the 99th percentile as an upper bound.

For estimating the register we took into account pitch values from all dialogs of each speaker, limited to utterances which were at least 1.5 s long, and which did not contain laughter. We then defined low and mid register of each speaker as follows. We took the 1st percentile of a speaker's pitch values as the reference minimum[2] and set the boundary between low and mid register to 3 semitones above that minimum. A range of 3 semitones has been argued to be the effective just-noticeable difference for pitch movements in communicative situations [22]. The range of the mid register was then also set to 3 semitones, calculated from that lower boundary. The high register then contains all values above those contained in the mid register.

Once we had determined the register boundaries for each speaker, we used them to classify turn transitions by determining the registers of the turns before and after the transition point. For the turn preceding the transition, we took one second of speech from the end of the turn, omitting the last 0.25 s to avoid possibly extreme values due to phrase boundary tones, and classified the turn as belonging to the register in which the median fell. For the normalizing approach, we used the speaker's own register boundaries in doing so. For the initializing approach, we used the following speaker's register boundaries. For the turn after the transition, we took one second of speech from the start of the turn, omitting the first 0.25 s to avoid possible initial phrase boundaries, and classified the turn as belonging to the register in which the median fell, always using the second speaker's own register boundaries.

### 2.4. Analysis: evaluating normalizing versus initializing approaches

A direct comparison between the normalizing and initializing approaches is difficult, since there is currently no perceptual data available with which to evaluate them. However, one way of testing the two approaches is to investigate correlations with other factors in the corpora. In this case, we investigated the degree to which pitch matching in both categorizing methods correlated with known aspects of the conversational situation, namely:

- Participant ratings of their interlocutor(s)
- Modality ("multimodal" face-to-face conversations versus "monomodal" blind conversations)
- Syntactic form (question-answer pairs versus statement-statement pairs)

Prosodic alignment or accommodation has often been argued to reflect positive relationship between speakers (e.g. [4]) or to arise when problematic aspects in conversation need to be smoothed over (e.g. [16]). On this basis, we expect that the participants' ratings of their interlocutors may correlate with the amount of pitch matching in the turn pairs. If pitch matching reflects a positive relationship, ratings for e.g. *friendly conversation* or *likeability* may correlate with increased pitch matching. Conversely, if pitch matching is used in contexts where the conversation is somehow problematic, *stressful conversation* or *problems* may correlate with pitch matching.

We predict that the *modality* of the conversation will impact prosodic performance, since participants who are face-to-face have more interactional resources available to them (e.g. eye contact, gesture). If prosodic matching is an intentional activity that speakers use to signal alignment with an interlocutor, then we may observe more matching in the "mono" conversations, when speakers do not have access to many non-verbal cues.

We also predict that the *syntactic form* of the first turn in each pair will also influence the degree to which matching pitch is found. Our methodology excludes the final high boundary

[1] Speakers were separated by a transparent screen in this condition for optimal signal separation.

[2] This was because the absolute minimum was more susceptible to outliers.

Table 1: *Poisson regression for "friendly conversation" in initializing data.* $R^2$*=0.462.*

Formula: MatchedPitch ˜ friendlyconv * modality + (1+friendlyconv | rater) + (1| ratee)

| Predictor | Estimate | SE | Z-value (p-value) |
|---|---|---|---|
| (intercept) (mono) | 0.675 | 0.664 | 1.02 (.3093) |
| Friendly | 0.695 | 0.348 | 2.00 (.0458*) |
| Modality (multi) | 0.701 | 0.306 | 2.29 (.0222*) |
| Friend*Mod | −0.631 | 0.170 | -3.72 (.0002*) |

Table 2: *Poisson regression for "friendly conversation" in normalizing data.* $R^2$*=0.444.*

Formula: MatchedPitch ˜ friendlyconv * modality + (1+friendlyconv | rater) + (1| ratee)

| Predictor | Estimate | SE | Z-value (p-value) |
|---|---|---|---|
| (intercept) (mono) | 0.568 | 0.784 | 0.72 (.4691) |
| Friendly | 0.717 | 0.385 | 1.86 (.0623) |
| Modality (multi) | 0.559 | 0.297 | 1.88 (.0597) |
| Friend*Mod | −0.438 | 0.167 | -2.62 (.0087*) |

tones which are often associated with questions, but is sensitive to the fact that (certain types of) questions, whether ending in a high boundary tone or not, often do not undergo declination to the same extent as declarative utterances (e.g. [23] for Dutch, a language that is typologically very similar to German). Thus we expect that question-statement turn pairs should be more likely to be exhibit high-mid or high-low patterns, while statement-statement pairs should be more likely to exhibit low-low or low-mid patterns (for example).

# 3. Results

Statistical models were calculated using the R statistical platform [24] with packages lme4 [25] and gmodels [26]; statistical significance was evaluated at $\alpha = .05$.

In the following, the term *turn pair* refers to any set of two turns at a turn transition (as identified in Section 2.2), whether statement-statement or question-answer. The term *pitch pair* refers to the assigned pitch values for a turn pair, e.g. low-low or high-mid. *Pitch matching* refers to cases in which the pitch categorization for both turns in a turn pair is the same: low-low, mid-mid, or high-high, with *pitch mismatching* referring to cases in which the pitch categorizations for the adjacent turns are different.

## 3.1. Interlocutor ratings

As described in Section 2.1, speakers rated the conversation and their interlocutor after each interaction. We carried out Poisson regressions to see if these ratings were able to predict matching behavior. We found that for the rating scale *friendly conversation* in the initializing data, cf. Table 1, participants' ratings predicted how often they themselves matched pitch with their interlocutor, in interaction with the modality of the conversation. That is, participants who afterwards gave positive ratings on this scale were participants who during the conversation also tended to produce pitch which matched their interlocutor's pitch in a preceding adjacent turn; given the significant interaction term, this effect was stronger in the monomodal conversations than in the multimodal conversations. In the normalizing data (Table 2), only the interaction term achieves statistical significance. The power model using the initializing categorization is a somewhat better fit to the data ($R^2$ of 0.462 compared to $R^2$ of 0.444 for the normalizing model), as well as giving a more easily interpreted result. Similar tests on the other proposed related ratings *likeability, stressful conversation, problems* did not achieve statistical significance.

## 3.2. Modality

Chi-squared tests on both the normalizing ($\chi^2$(8, N=1518) = 33.75, p<.001) and initializing data sets ($\chi^2$(8, N=1518) = 33.79, p<.001) attain statistical significance for the distribution of pitch matching and mismatching. The models derived using the two pitch categorization methods are nearly identical in this case. Investigating the residuals for each square of the models (cf. Table 3), we find that low-low pitch pairs occur significantly more often than predicted in multimodal conversation, and significantly less often than predicted in monomodal conversation.

Table 3: *Chi-squared distribution for pitch pairs by modality (initializing data); the table contains the number of observations in each condition (monomodal and multimodal), with residual values indicating the extent to which the number of observations differs from a predicted non-significant distribution. Observations with residuals larger than +/-2 contribute significantly to the model.*

| Pitch Pair | Mono Obs. (Resid.) | Multi Obs. (Resid.) |
|---|---|---|
| high-high | 53 (-0.645) | 65 (0.634) |
| high-low | 9 (-1.492) | 21 (1.464) |
| high-mid | 85 (0.010) | 88 (-0.010) |
| low-high | 38 (0.985) | 28 (-0.967) |
| low-low | 34 (-2.152*) | 66 (2.113*) |
| low-mid | 141 (-0.903) | 169 (0.887) |
| mid-high | 83 (1.659) | 58 (-1.629) |
| mid-low | 30 (-1.547) | 51 (1.519) |
| mid-mid | 272 (1.732) | 227 (-1.700) |

## 3.3. Questions and statements

As predicted, certain types of pitch pairs—specifically certain types of pitch mismatches—were more common in either question-answer turn pairs or statement-statement turn pairs. Chi-squared tests on both the normalizing ($\chi^2$(8, N=1518) = 63.33, p<.001; cf. Table 4) and initializing data sets ($\chi^2$(8, N=1518) = 47.23, p<.001) attain statistical significance for the different distribution of pitch pairs in question-answer and statement-statement turn pairs. It appears that the normalizing approach gives a slightly more differentiated picture of the use of prosody in questions vs. statements, as the differences between question-answer pairs and statement-statement pairs are more pronounced when assessing register using the normalizing method: there are more deviations from the overall distribution of transition types.

One observation that is consistent across both of these chi-squared models is that question-answer pairs are relatively unlikely to have low-mid or low-high pitch realization. This is consistent with literature suggesting that interrogatives in many forms tend to have less or no pitch declination compared to

Table 4: *Chi-squared distribution for pitch pairs by syntactic structure (normalizing data); the table contains the number of observations in each condition (question-answer and statement-statement), with residual values indicating the extent to which the number of observations differs from a predicted non-significant distribution. Observations with residuals larger than +/-2 contribute significantly to the model.*

| Pitch Pair | Qu-An Obs. (Resid.) | St-St Obs. (Resid.) |
|---|---|---|
| high-high | 14 (1.297) | 39 (-0.622) |
| high-low | 7 (0.867) | 20 (-0.416) |
| high-mid | 55 (4.778*) | 101 (-2.292*) |
| low-high | 6 (-2.072*) | 67 (0.994) |
| low-low | 10 (-0.901) | 61 (0.432) |
| low-mid | 19 (-4.506*) | 255 (2.162*) |
| mid-high | 40 (0.454) | 159 (-0.218) |
| mid-low | 23 (0.404) | 90 (-0.194) |
| mid-mid | 110 (0.662) | 442 (-0.318) |

declaratives. This raises the question of whether there is any consistency in other formal aspects of question-answer turn pairs which are realized with these relatively infrequent pitch patterns. An inspection of the individual turn pairs revealed that, after discarding two items which had been incorrectly classified due to octave errors and two which had overlap earlier in the turn (which the algorithm described above could not detect), all of the turns classified as low-mid or low-high by either the normalizing or initializing methods had one or both of the following characteristics: (1) the question was a tag question, in many cases *oder?* "right?", and/or (2) the question ended with a very high boundary tone[3]. Statistical models were not able to successfully predict the presence/absence of *oder* compared to a different tag question or the presence of a high boundary tone on the basis of the categorizations, but this may be due to the relatively small number of turns involved (N=21 for normalizing, N=34 for initializing).

# 4. Discussion

## 4.1. Normalizing versus initializing pitch approaches

Our data have given somewhat conflicting evidence for the two pitch classification approaches used. In the case of the participant ratings, there was a slightly stronger correlation between participants' ratings of their interlocutor's friendliness and pitch matching as calculated using the initializing approach. However, in terms of identifying modality, there is no substantial difference between the two pitch evaluation methods, and in investigating pitch mismatching in question-answer and statement-statement pairs, the model using the normalizing approach is a somewhat better fit to the data (i.e. the $\chi^2$ value of the model using the normalizing classification is higher, although both models attain statistical significance).

While we therefore cannot make an immediate claim as to which approach (i.e. normalizing versus initializing) to categorizing pitch is more accurate, the distribution of the results suggests that our conclusions about whether normalizing or initializing is a more appropriate way of addressing the data may depend on what kinds of phenomena we investigate. The "linguistic" contrast of statements and questions appears to be more closely represented by the normalizing data, while the "paralinguistic" factor of conversational participants' attitudes about the conversation is more closely approximated by the initializing data. At this stage it is impossible to know whether one of these should be prioritized above the other in such investigations. To this end, we are preparing a perception study to test listeners' assessments of the different pitch pairings.

## 4.2. Participant ratings

While it is often proposed that prosodic accommodation, such as pitch matching, is associated with positive feelings between interlocutors, our data provide particular insight into this question in that it is a rater's own decision to match pitch with an interlocutor, rather than the interlocutor's pitch matching, that is correlated with positive assessments on the *friendly conversation* scale, at least in the monomodal conversations. Although it is impossible to determine causality in this case, it appears that raters evaluated the friendliness of the conversation on the basis of their own friendliness (assuming pitch matching is in fact related to friendliness) rather than that of their interlocutor.

## 4.3. Modality and use of prosody

Our data show significant differences of turn pitch realization across the two modalities tested (i.e. multimodal versus monomodal, cf. Section 2.1). Specifically, low-low pairs were more frequent in the multimodal condition, and less frequent in the monomodal condition. Participant ratings of their interlocutors were also correlated with pitch matching, but with a stronger effect in the monomodal than in the multimodal conversations. Furthermore, no differences were found between the modalities in speakers' ratings of their conversations overall, or in structural differences in number of question-answer versus statement-statement turn pairs. This supports the interpretation that conversational participants were deploying modality-specific prosodic strategies. When they had access to other information such as facial expression, eye gaze, body movements/gesture, etc., there was apparently less need to be prosodically expressive. This finding demonstrates the importance of comparing data from different modalities when investigating the role of prosodic variation in expressing meaning.

# 5. Conclusions

Using two methods of pitch classification, based on normalizing and initializing approaches to pitch, we investigate pitch relationships between adjacent turns in German conversation. We find that conversational modality, syntactic structure, and participant attitudes are all related to the ways in which conversationalists deploy their pitch resources. Furthermore, we show that the normalizing and initializing approaches may highlight different aspects of pitch use in conversation; a planned perception experiment will shed further light on this issue. Such future research must take into account the interpretive aspects of its methodology of pitch measurement when investigating pitch patterns in conversational speech.

# 6. Acknowledgements

This research was supported by the SFB 732 *Incremental Specification in Context* (University of Stuttgart), funded by the German Research Foundation (DFG).

[3]Recall that boundary tones were functionally excluded from our pitch analysis since our algorithm did not include the final 250ms of a turn. Thus the high boundary tones here were preceded by longer stretches of particularly low pitch.

# 7. References


[1] J. J. Gumperz, "Contextualization and understanding," in *Rethinking Context: Language as an Interactive Phenomenon*, A. Duranti and C. Goodwin, Eds. Cambridge: Cambridge University Press, 1992, pp. 229–252.

[2] E. A. Schegloff, "Reflections on quantification in the study of conversation," *Research on Language and Social Interaction*, vol. 26, pp. 99–128, 1993.

[3] ——, "Accounts of conduct in interaction: Interruption, overlap and turn-taking," in *Handbook of Sociological Theory*, J. H. Turner, Ed. New York: Plenum, 2002, p. 287321.

[4] E. Couper-Kuhlen and M. Selting, "Towards an interactional perspective on prosody and a prosodic perspective on interaction," in *Prosody in Conversation: Interactional Studies*, E. Couper-Kuhlen and M. Selting, Eds. Cambridge: Cambridge University Press, 1996, pp. 11–56.

[5] R. Ogden, "Phonetics and social action in agreements and disagreements," *Journal of Pragmatics*, vol. 38, pp. 1752–1775, 2006.

[6] C. Gallois and V. J. Callan, "Communication accommodation and the prototypical speaker: Predicting evaluations of status and solidarity," *Language & Communication*, vol. 8, 1988.

[7] H. Giles, J. Coupland, and N. Coupland, *Contexts of Accommodation*. Cambridge, UK: Cambridge University Press, 1991.

[8] M. J. Pickering and S. Garrod, "The interactive-alignment model: Developments and refinements," *Behavioral and Brain Sciences*, vol. 27, 2004.

[9] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell, "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction," *Speech Communication*, vol. 58, pp. 11–34, 2014.

[10] D. N. Honorof and D. H. Whalen, "Perception of pitch location within a speaker's F0 range," *Journal of the Acoustical Society of America*, vol. 117, 2005.

[11] J. Bishop and P. Keating, "Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 1100–1112, 2012.

[12] J. Kuang and M. Liberman, "Pitch-range perception: the dynamic interaction between voice quality and fundamental frequency," *Proceedings of 17th Interspeech, San Francisco, CA, USA*, pp. 1350–1354, 2016.

[13] D. R. Ladd, *Intonational Phonology*. Cambridge, UK: Cambridge University Press, 2008.

[14] J. Gorisch, B. Wells, and G. J. Brown, "Pitch contour matching and interactional alignment across turns: An acoustic investigation," *Language & Speech*, vol. 55, pp. 57–76, 2012.

[15] M. A. Sicoli, T. Stivers, N. J. Enfield, and S. C. Levinson, "Marked initial pitch in questions signals marked communicative function," *Language & Speech*, vol. 58, pp. 204–223, 2014.

[16] M. Zellers and R. Ogden, "Exploring interactional features with prosodic patterns," *Language & Speech*, vol. 57, no. 3, pp. 285–309, 2014.

[17] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech," *Proceedings of 14th Interspeech, Lyon, France*, pp. 525–529, 2013.

[18] ——, "Social factors in convergence of f1 and f2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production, Cologne*, 2014.

[19] R. O. Sikveland and D. Zeitlyn, "Using prosodic cues to identify dialogue acts: methodological challenges," *Text & Talk*, vol. 37, no. 3, pp. 311–334, 2017.

[20] D. Hirst, "A Praat plugin for MOMEL and INTSINT with improved algorithms for modelling and coding intonation," in *Proceedings of ICPhS XVI*, 2007, pp. 1233–1236.

[21] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer [computer program]," 2017, version Praat 6.0.24. [Online]. Available: http://www.praat.org/

[22] J. 't Hart, "Differential sensitivity to pitch distance, particularly in speech," *Journal of the Acoustical Society of America*, vol. 69, no. 1, pp. 811–821, March 1981.

[23] J. Haan, V. J. van Heuven, J. Pacilly, and R. van Bezooijen, "An anatomy of Dutch question intonation," *Linguistics in the Netherlands*, vol. 14, no. 1, pp. 97–108, 1997.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: https://www.R-project.org/

[25] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[26] G. R. Warnes, B. Bolker, T. Lumley, and R. C. Johnson, *gmodels: Various R Programming Tools for Model Fitting*, 2015. [Online]. Available: https://CRAN.R-project.org/package=gmodels