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

To see or not to see: Interlocutor visibility and likeability influence convergence in intonation

Katrin Schweitzer, Michael Walsh, Antje Schweitzer

Institute for Natural Language Processing, University of Stuttgart, Germany
Katrin.Schweitzer/Michael.Walsh/Antje.Schweitzer@ims.uni-stuttgart.de

Abstract

In this paper we look at convergence and divergence in intonation in the context of social qualities. Specifically we examine pitch accent realisations in the GECO corpus of German conversations. Pitch accents are represented as 6-dimensional vectors where each dimension corresponds to a characteristic of the accent's shape. Convergence/divergence is then measured by calculating the distance between pitch accent realisations of conversational partners. A decrease of distance values over time indicates convergence, an increase divergence. The corpus comprises dialogue sessions in two modalities: partners either saw each other during the conversation or not. Linear mixed model analyses show convergence as well as divergence effects in the realisations of H*L accents. This convergence/divergence is strongly related to the modality and to how much speakers like their partners: generally, seeing the partner comes with divergence, whereas when the dialogue partners cannot see each other, there is convergence. The effect varies, however, depending on the extent to which a speaker likes their partner. Less liking entails a greater change in the realisations over time – stronger divergence when partners could see each other, and stronger convergence when they could not.
Index Terms: convergence, accommodation, social factors, pitch accents, intonation, similarity, distance

1. Introduction

Phonetic convergence and divergence are dynamic processes by which speakers adapt aspects of their speech towards, or away from, the speech of their dialogue partner. Proponents of Communication Accommodation Theory (CAT, e.g. [1, 2, 3, 4]) argue that convergence is symptomatic of a speaker's (unconsious) desire for social inclusion. Divergence, in contrast, reflects the need to distance oneself from a speech partner's social group. Additional reasons to adapt one's speech are to increase intelligibility and efficiency of communication [5, 6, 7]. Thus, social factors and the communication setting are clearly important when exploring convergence/divergence.

Investigations into these phenomena have been carried out with regard to a number of speech properties including speech rate [8, 9, 10, 11, 12], voice onset time [13, 14], vowel accommodation [15, 16, 17], pitch [18, 11, 19, 20, 21], language distance [22] and frequency of occurrence effects [12], among others. To date, however, to our knowledge, the extent to which interlocutors might adapt their pitch accent realisations in a conversational dialogue has not been explored. In this paper, we examine exactly this phenomenon in a corpus of conversational German where dialogues took place under two modalities: one where interlocutors could see each other (*visible* modality) and one where they could not (*blind* modality). Furthermore, in both modalities, social ratings of likeability were provided by the interlocutors after the dialogue. We hypothesise that over the course of the dialogues interlocutors will converge in their pitch accent realisations in the contexts where they find their dialogue partner likeable. In addition, we anticipate that the interlocutors who could see each other will be more prone to converge. It is important to note that while these hypotheses seem intuitive there is considerable inconsistency in the literature with regard to phonetic convergence. For example, while convergence with respect to speech rate over complete dialogues has been reported [11], the effects fall below significance when examined at the turn level, and an unexpected general effect of divergence when looking at speech rate at the turn level has also been reported [8], but this was countered in the same study by convergence where participants liked each other. Furthermore, although enhanced convergence in the *visible* modality in conversational speech has been observed [23], more recent work examining convergence where partners could see each other in a speech shadowing task under noisy/clean auditory conditions failed to duplicate the convergence finding in the clean speech condition [24].

2. The GECO Database

The **Ge**rman **Co**nversations (GECO) speech database [25] consists of spontaneous conversations between previously unacquainted female German speakers on topics of their choice. Each dialogue lasted approx. 25 minutes. There are 46 dialogues (approx. 21 hours of dialogue) in the database. The recordings were automatically annotated on the segment, syllable, word, and prosodic levels, including estimations of GToBI(S) pitch accent labels (Stuttgart variety of German ToBI, cf. [26]). Due to a lack of a) training data for spontaneous German speech and b) reliable POS-tagging for non-standard data, the automatic pitch accent annotations provided with the database do not reach the standards of state-of-the-art prosodic annotation as e.g. provided in [27]. But they were sufficient for our purposes, since we only employed them to split up the data coarsely into one dataset consisting of rising (L*H) and one of falling (H*L) pitch accents.

The dialogues were carried out in two different settings. In one modality speakers were separated by a sound-treated wall and could not see each other (*blind* modality). After their conversation, they rated each other in terms of likeability by filling in a questionnaire. In the other modality they could see each other through a transparent screen (*visible*), and also rated each other in terms of likeability once the conversation was over.

Likeability was captured by four items (likeable, friendly, socially attractive and relaxed), with a 5-point Likert scale for each.[1] In the database the Likert scale answers were transformed to integers from -2 to +2. We added the values of the four likeability items to obtain composite scores. Even though

[1]For other social aspects available and details concerning the items see [8, 25].

http://dx.doi.org/10.21437/Interspeech.2017-1248

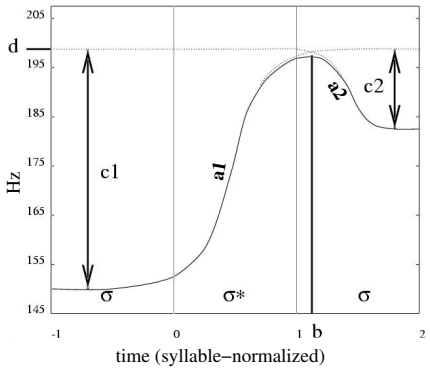

Figure 1: *The PaIntE model function and its parameters. Figure adapted from [28].*

negative scores are rare, the composite scores exhibit reasonable variation (ranging from -2 to 8).

3. The PaIntE Model

The PaIntE model of intonation captures the shape of rises, falls, and peaks in the F_0-contour by approximating the smoothed curve with the help of a model function. This function operates over time on a three-syllable window. The function term contains 6 free parameters which are estimated in such a way that they fit the respective stretch of the F_0-contour best. The six parameters correspond to linguistically interpretable characteristics of pitch accent shape. Specifically, parameter d denotes the absolute height of the peak in Hz, parameter b the temporal alignment of the peak within the 3-syllable window (normalised over syllable time), parameters c1 and c2 the size of the rise and fall before and after the peak, and a1 and a2 correspond to the steepness of the rise/fall.

Figure 1 illustrates the parameters. They are calculated over the span of the accented syllable ($\sigma*$) and its immediate neighbours. The x-axis indicates time (normalised for syllable duration, i.e. the syllable bearing the accent spans from 0 to 1) and the y-axis displays the fundamental frequency in Hertz.

PaIntE usually models the peak with the help of two sigmoids, one modelling the rise before the peak, the other modelling the fall after. In some cases, however, modelling with one sigmoid provides a better fit. In these cases, the unused parameter denoting the steepness of the rise (or fall) is not used in our experiments. If the modelling outcome is still not a good fit to the contour, the model just returns the mean F_0 value.

4. Methodology

4.1. Calculation of similarity

From the GECO database, we extracted all syllables that received a pitch accent label in the automatic annotation and that were modelled with at least one sigmoid of the PaIntE function. We removed outlying pitch accents where the value in one or more of the PaIntE dimensions lay below the 1st or above the 99th percentile for a given speaker. We chose these boundaries in order to retain productions that are atypical for a given speaker, and hence might result from accommodation, while at the same time factoring out data that is likely to be erroneous. Visual inspection of the density plots for each dimension confirmed that this was a reasonable approach. The values of each PaIntE dimension were then normalised on a speaker basis by z-scoring.

Based on the automatic pitch accent annotations provided with the GECO database (cf. section 2), we created two datasets for the two most frequent accent types, one for H*L accents (16484 tokens) and one for L*H accents (19419 tokens).

We carried out the following procedure for both datasets separately: For each turn (specified as a continuous passage of speech by one speaker) in a dialogue, we extracted all pitch accents of the given type. If more than one accent of this type was present, we calculated a centroid for that type for that turn. In this way, every turn had a single representative pitch accent token/centroid, for the given type. We term this accent the turn-accent. The H*L dataset comprised 10814 turn-accents, the L*H dataset 11297 turn-accents.

Over the course of the dialogue, we then calculated the Euclidean distance of each turn-accent to the partner's preceding turn-accent as a measure of their similarity/dissimilarity. In cases where the preceeding turn had no accent (of the given type), no comparison was made. This resulted in 4769 distance values for H*L turn-accents and 4564 for L*H turn-accents.

Note that this similarity calculation gives insight into convergence/divergence effects over the course of a dialogue, but it does not target the question of who is converging to whom.

4.2. Statistical analysis

To investigate possible effects of convergence/divergence in the context of likeability, we fitted a linear mixed model for each turn-accent dataset. The models predict Euclidean distance between turn-accents as a function of the time of their realisation in the dialogue, the scores for likeability, and the modality of the dialogues. In order to avoid multicollinearity, we centred the likeability scores and the time in dialogue. The dialogue time was also z-scored on a dialogue-basis. Specifically, the fixed factors then were: time (*c.time*), likeability (*c.like*), and modality (*mod*, categorical, with levels *visible* and *blind*, the contrast was sum-coded in the models), as well as all interactions of these three factors. To control for effects of specific dialogue partners we added a random intercept for *partners* and by-*partners* random slopes for *c.time*.

Significance of a factor in a model was determined by performing likelihood ratio model comparisons between that model and a model with the factor in question removed (cf. [29, 29, 30]), using the *afex*-package [31] in R 3.3.2 [32]. The significance level was $\alpha = 0.05$.

5. Results

5.1. Convergence/divergence and likeability in the H*L dataset

The model term for the model predicting distance of H*L turn-accents by time, modality and likeability scores is given in (1).

$$\begin{aligned} dist \sim\ & c.time * c.like * mod \\ & + (1|partners) + (0 + c.time|partners) \end{aligned} \tag{1}$$

The variance inflation factor for all predictors was uncritical (VIF<2), even under conservative criteria [33], as was the condition number (κ <4), hence multicollinearity was not a problem in the model. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

Table 1: *Coefficients, Standard Errors, and t-values of the Model predicting distance by time, likeability and modality. Last column gives the p-values resulting from the likelihood ratio comparisons.*

	Estimate	Std. Error	t value	$p(\chi^2)$	
(Intercept)	2.774	0.034	82.182		
c.time	-0.009	0.021	-0.435	0.666	
c.like	-0.011	0.011	-1.03	0.307	
mod1	-0.094	0.026	-3.601	0.0003	***
c.time:c.like	-0.012	0.009	-1.286	0.203	
c.time:mod1	-0.029	0.02	-1.417	0.161	
c.like:mod1	0.002	0.011	0.2	0.844	
c.time:c.like:mod1	0.024	0.009	2.701	0.008	**

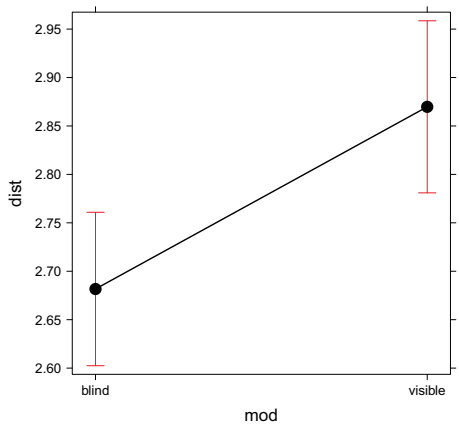

Figure 2: *Effect of* modality *on* distance *values in the model predicting distance as a function of time, likeability and session for H*L accents. Confidence level is 0.95.*

In the model comparisons, modality (*mod*) was a significant factor in predicting distance (*dist*) with $p < 0.001$, as was the three-way interaction of time, likebility and modality (*c.time:c.like:mod*, $p < 0.01$). Table 1 gives an overview of the results; p-values below the significance level of $\alpha = 0.05$ are indicated.

The coefficient for *mod1* is -0.09. Since the categorical variable *mod* is sum coded, the coefficient corresponds to the difference of the distance value of the first level of *modality*, i.e. the condition *blind*, to the grand mean. There are two modalities, i.e. two levels of the variable *mod*, which means that the difference in distance values between *visible* and *blind* sessions is 0.18. This effect calculated by the model is visualised in figure 2 using the *effects*-package [34] in R [32].

Figure 3 presents the complex relationship of the three-way interaction of *c.time*, *c.like*, and *mod*, as predicted by the model (again, visualised using the *effects* package [34]). The solid black line shows the effect predicted for *blind* sessions, the dashed red line for *visible* sessions. Normalised and centered time in the dialogue is on the x-axis of each panel; *distance* is shown on the y-axis. The three panels show the effect for values of least, mid and highest liking on the centered likeability scale. As is visible from the plot, in the *blind* condition, distance values are predicted to decrease over time in the case of least liking, this decrease is less pronounced for mid liking, and completely flattened out for highest liking scores, where the distance between the partners' realisations stays more or less constant throughout the dialogue. In the *visible* condition however, partners' realisations are predicted to become more distant over time in the case of least liking (the dashed red line shows the increase in distance values over the course of the dialogue), the increase in distance values is less pronounced for mid liking, and in the case of highest liking scores, the distance values are predicted to decrease somewhat.

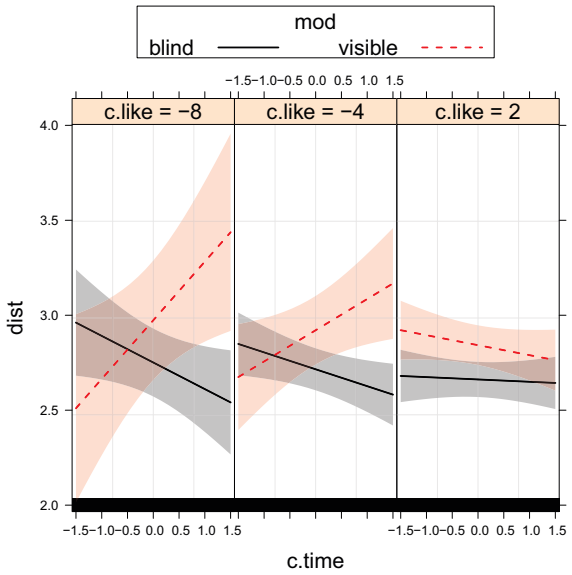

Figure 3: *Effect of the three way interaction* c.time:c.like.c.mod *on* distance *values in the model predicting distance as a function of time, likeability and session for H*L accents. The shaded areas correspond to estimated confidence intervals ($\alpha = .95$). The values of the likeability scores in the three panels (from left to right) correspond to least, mid, and highest liking on the centered likeability scale.*

5.2. Convergence/divergence and likeability in the L*H dataset

The model predicting the distance of the interlocutors' L*H realisations yielded no significant results.

6. Discussion

The results presented in section 5 indicate a complex relationship between pitch accent distances, the time in a dialogue, the modality of the dialogue, and the extent to which the interlocutors like each other.

However, it has to be noted that we only found significant effects of convergence and divergence for accents which were classfied as H*L by the automatic classfier. The absence of an effect for L*H accents might indicate that those accents are realised more consistently across speakers and contexts with respect to their tonal quality. It is of course possible that there is still convergence in realisation of these accents on non-tonal levels. That no effect could be observed for L*H could also be due to a more technical reason: since the PaIntE model is designed to model a peak, the exact shape of L*H accents is not captured as well as is the case for H*L tokens. The function operates over a 3-syllable window around the syllable bearing

the accent. Therefore, for a canonical L*H, the peak would be expected in the third syllable of the analysis window, and consequently, some characteristics of the fall, which could well be realised even later in the utterance, might be missing in the characterisation of the F_0-contour provided by the PaIntE model.

In the following we will discuss the significant results for H*L and how they relate to our hypotheses.

6.1. Distance/similarity in the two modalities

We expected greater convergence of the turn-accents in the *visible* modality, in line with [23]. This expectation was not met. If convergence is defined as a reduction of distance values over time, an overall effect of *modality* on convergence should have resulted in a significant effect for an interaction of time and modality. However, we found a main effect for modality alone. This effect (cf. figure 2) shows that partners' realisations are generally expected to be more distant from each other in the *visible* condition, i.e. when interlocutors were able to see each other. Conversely, when the partners could not see each other, their realisations were more similar in general, irrespective of the time in the dialogue. This might be due to the fact that in the *blind* condition speakers had greater cognitive resources available to them for speech monitoring as no decoding of visual information is necessary (cf. [20]). Consequently, speaker productions are less variable, and, as a result, there is less variability overall. Moreover, in the *visible* condition, speakers use non-verbal communication such as (facial) gestures, body language etc., resulting in more physical movement and possibly introducing yet another source of within-speaker variability, introducing more chance for across-speaker variability.

6.2. Convergence/divergence and social factors in different modalities

We expected greater convergence for conversations where speakers ranked their partners high on the likeability scale [17, 1, 2, 3]. The results of our analyses, however, show that the relationship between convergence/divergence, the modality of the session and the social qualities attributed to the conversational partner, is a complex one.

The expectation that greater liking results in greater convergence was only met in the *visible* modality. The trajectory of the dashed red line in figure 3 shows the predicted degree of divergence/convergence. In the case of low likeability scores (left panel), the line rises. That is, over the course of the dialogue (from left to right in the left panel) the distances between H*L turn-accent realisations is expected to become larger – the partners diverge. For medium likeability scores (middle panel) the line still rises but the divergence is less pronounced. In the case of the highest likeability scores (right panel), the line falls over the course of the dialogue, indicating that the H*L realisations are expected to be more similar to each other by the end of the dialogue than they were in the beginning. Hence, for the *visible* condition, the model predicts divergence in the case of low likeability which becomes smaller and turns into slight convergence as likeability scores increase.

In the *blind* modality, however, the picture is reversed. In the case of low likeability scores (the solid black line in fig. 3) H*L turn-accent realisations are predicted to become less distant over time: interlocutors converge. The convergence effect is still present in the middle panel (with medium likeability scores), though less pronounced. However, the line is almost flat in the rightmost panel, which displays the predicted trajectory of the distance values over the course of the dialogue where partners rated each other highly. The distance between realisations is almost constant.

As mentioned above, the results for the *visible* modality are in keeping with our initial hypotheses and the general expectations that mutual liking entails convergence [17, 1, 2, 3]: the model predicts less divergence (i.e. more convergence) with increasing likeability scores. The convergence effect in the *blind* condition is unexpected, but to some extent in line with the assumption that convergence can serve to increase intelligibility and efficiency in communication [5, 6, 7]. That is, the absence of information available from the visual channel could increase the need for convergence, especially when the initial realisations are distant (as happens in the cases of low likeability scores, see left panel of fig. 3) and when the dialogue partners are aiming to have a "successful" conversation (as is likely since they both agreed to take part in conversation recordings). In the cases of greater likeability scores, the H*L realisations are already less distant at the beginning of the dialogue (right panel fig. 3) potentially making convergence less necessary.

Note that it is not clear from the data whether the low likeabilty scores result from the effort required to adapt (unconsciously affecting how partners view each other), the initial distance between the interlocutors productions (partners unconsciously perceive greater social distance), or from the contents of the conversations. If low likeability scores co-occur with conversations where the content was to some extent controversial or uncomfortable for one or both speakers, the effect observed here is similar to the one reported in a study that finds prosodic features to match, when "problematic aspects" of a conversation needed to be smoothed over [35]. In the *visible* modality, interlocutors have more means to resolve difficult conversation points (e.g. by making eye contact, smiling, etc.).

7. Conclusion

We examined pitch accent convergence in natural conversational speech of female German speakers in the context of likeability and two conversation modalities. Generally, we found the realisations of H*L turn-accents to be more different between the two partners when they could see each other, compared to cases where verbal communication was the only communciation channel. Moreover, when partners could see each other, we found decreased divergence with increasing liking. While this result met our expectations, the results from dialogues where speakers could not see each other were surprising: we observed more convergence in low likeability contexts, potentially to compensate for missing (visual) information, and increased initial distance in partners' productions, in a situation where a successful conversation was the goal.

These results present various options for future work, e.g. examining similarity in perceptual terms. Furthermore, as one of the reviewers pointed out, it might be possible that the communicative role of pitch accents influences the degree to which they are subject to convergence (e.g. an accent marking contrast might be less conducive to convergence). The function of pitch accents however is nuanced, and even results on read speech are not clear-cut [36, 37]. Therefore, a more fine grained analysis of the communicative function of the accents in the corpus can shed further light on the mechanisms behind convergence.

8. Acknowledgements

This work was funded by the German Research Association DFG (SFB 732, projects INF, A2, and A4).

9. References

[1] H. Giles and P. M. Smith, "Accommodation theory: Optimal levels of convergence," in *Language and Social Psychology*, H. Giles and R. St. Clair, Eds. Oxford: Blackwell, 1979, pp. 45–65.

[2] H. Giles, N. Coupland, and J. Coupland, "Accommodation theory: Communication, context and consequence," in *Contexts of Accommodation*, H. Giles, N. Coupland, and J. Coupland, Eds. Cambridge University Press, 1991, pp. 1–68.

[3] C. A. Shepard, H. Giles, and B. A. Le Poire, "Communication Accommodation Theory," in *The New Handbook of Language and Social Psychology*, W. P. Robinson and H. Giles, Eds. John Wiley & Sons, 2001, pp. 33–78.

[4] H. Giles and T. Ogay, "Communication accommodation theory," in *Explaining communication: Contemporary theories and exemplars*, B. Whaley and W. Samter, Eds. Mahwah, NJ: Lawrence Erlbaum, 2006, pp. 293–310.

[5] H. C. Triandis, "Cognitive similarity and communication in a dyad," *Human Relations*, vol. 13, pp. 175 – 183, 1960.

[6] M. Natale, "Social desirability as related to convergence of temporal speech patterns," *Perceptual and Motor Skills*, vol. 40, pp. 827–830, 1975.

[7] C. Gallois, H. Giles, E. Jones, A. C. Cargile, and H. Ota, "Accommodating intercultural encounters: Elaborations and extensions," in *Intercultural communication theory*, R. Wiseman, Ed. Thousand Oaks, CA: Sage, 1995, pp. 115 – 147.

[8] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013, Lyon)*, 2013, pp. 525–529.

[9] J. T. Webb, "Interview synchrony: an investigation of two speech rate measures in an automated standardized interview," in *Studies in dyadic communication*, A. Siegman and B. Pope, Eds. Pergamon Press, 1972, pp. 115–133.

[10] R. Street, N. J. Street, and A. Van Kleek, "Speech convergence among talkative and reticent three year-olds," *Language Sciences*, vol. 5, no. 1, pp. 79–96, 1983.

[11] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, 2011, pp. 3081–3084.

[12] A. Schweitzer and M. Walsh, "Exemplar dynamics in phonetic convergence of speech rate," in *Proceedings of Interspeech 2016*, San Francisco, USA, 2016, pp. 2100–2104.

[13] K. Shockley, L. Sabadini, and C. A. Fowler, "Imitation in shadowing words," *Perception and Psychophysics*, vol. 66, no. 3, pp. 422–429, 2004.

[14] K. Nielsen, "Specificity and abstractness of vot imitation," *Journal of Phonetics*, vol. 39, no. 2, pp. 132–142, 2011.

[15] M. E. Babel, "Phonetic and social selectivity in speech accommodation," Ph.D. dissertation, University of California, Berkeley, 2009.

[16] M. Babel, "Dialect divergence and convergence in new zealand english," *Language in Society*, vol. 39, no. 04, pp. 437–456, 2010.

[17] J. S. Pardo, R. Gibbons, A. Suppes, and R. M. Krauss, "Phonetic convergence in college roommates," *Journal of Phonetics*, vol. 40, no. 1, pp. 190–197, 2012.

[18] J. S. Pardo, "Expressing oneself in conversational interaction," *Expressing oneself/expressing ones self: Communication, cognition, language, and identity*, pp. 183–196, 2010.

[19] S. Gregory and S. Webster, "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions," *Journal of Personality and Social Psychology*, vol. 70, no. 6, pp. 1231–1240, 1996.

[20] S. W. Gregory, B. E. Green, R. M. Carrothers, K. A. Dagan, and S. W. Webster, "Verifying the primacy of voice fundamental frequency in social status accommodation," *Language & Communication*, vol. 21, no. 1, pp. 37 –60, 2001.

[21] M. Babel and D. Bulatov, "The role of fundamental frequency in phonetic accommodation," *Language and Speech*, vol. 55, no. 2, pp. 231–248, 2012, pMID: 22783633. [Online]. Available: http://dx.doi.org/10.1177/0023830911417695

[22] M. Kim, W. S. Horton, and A. R. Bradlow, "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance," *Laboratory Phonology*, vol. 2, no. 1, pp. 125 – 156, 2011.

[23] J. W. Dias and L. D. Rosenblum, "Visual influences on interactive speech alignment," *Perception*, vol. 40, no. 12, pp. 1457–1466, 2011.

[24] ——, "Visual enhancement of alignment to noisy speech," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, p. 060139, 2013. [Online]. Available: http://asa.scitation.org/doi/abs/10.1121/1.4800689

[25] A. Schweitzer and N. Lewandowski, "Social factors in convergence of F1 and F2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production, Cologne*, 2014.

[26] J. Mayer, "Transcribing German intonation – the Stuttgart system," Universität Stuttgart, Tech. Rep., 1995. [Online]. Available: http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html

[27] A. Schweitzer, *Production and Perception of Prosodic Events – Evidence from Corpus-based Experiments*. Doctoral dissertation, Universität Stuttgart, 2010.

[28] G. Möhler, "Improvements of the PaIntE model for F_0 parametrization," Institute of Natural Language Processing, University of Stuttgart, Tech. Rep., 2001, draft version.

[29] H. Baayen, *Analyzing Linguistic Data*. Cambridge University Press, 2008.

[30] B. Winter, "Linear models and linear mixed effects models in r with linguistic applications." arXiv: 1308.5499, 2013. [Online]. Available: http://arxiv.org/pdf/1308.5499.pdf

[31] H. Singmann, B. Bolker, J. Westfall, and F. Aust, *afex: Analysis of Factorial Experiments*, 2016, r package version 0.16-1. [Online]. Available: https://CRAN.R-project.org/package=afex

[32] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[33] A. F. Zuur, E. N. Ieno, and C. S. Elphick, "A protocol for data exploration to avoid common statistical problems," *Methods in Ecology and Evolution*, vol. 1, no. 1, pp. 3–14, 2010. [Online]. Available: http://dx.doi.org/10.1111/j.2041-210X.2009.00001.x

[34] J. Fox, "Effect displays in r for generalised linear models," *Journal of Statistical Software*, vol. 8, no. 15, pp. 1–27, 2003. [Online]. Available: http://www.jstatsoft.org/v08/i15/

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

[36] C. Féry, "The Fallacy of Invariant Phonological Correlates of Information Structural Notions," in *The Notions of Information Structure*, C. Féry, G. Fanselow, and M. Krifka, Eds. Potsdam: Universitätsverlag Potsdam, 2007, vol. 6, pp. 161–204.

[37] K. Schweitzer, A. Riester, M. Walsh, and G. Dogil, "Pitch Accents and Information Status in a German Radio News Corpus," in *Proceedings of Interspeech 2009*, Brighton, UK, 2009, pp. 877–880.