Convergence of Articulation Rate in Spontaneous Speech

Antje Schweitzer, Natalie Lewandowski

Institute for Natural Language Processing (IMS), University of Stuttgart, Germany
{antje.schweitzer,natalie.lewandowski}@ims.uni-stuttgart.de

Abstract

This paper presents results from a project on phonetic convergence in German spontaneous speech. We used linear mixed models to examine 22 unimodal and 24 multimodal dialogs for articulation rate. We show that speakers' local articulation rates are influenced by the preceding rates of their interlocutors, and that the direction of this influence (i.e., divergence or convergence) depends on social factors, viz. interactants' mutual likeability scores. More specifically, we found that in general there was a "default" effect of divergence in articulation rates which was not mediated by social factors. However, this effect was weakened or reversed for higher mutual liking scores, i.e. the degree of convergence increased with the liking scores. Furthermore, while it has recently been suggested that convergence may be enhanced in multimodal settings, we did not find an effect of modality on convergence. However, there was an effect of modality on articulation rate in general.

Index Terms: phonetic convergence, accommodation, alignment, entrainment, articulation rate, social factors

1. Introduction

Convergence is the process of accommodating one's style of speech to that of an interlocutor in a way that it becomes more similar to the interlocutor's style. A related phenomenon is phonetic imitation, where speakers become more similar to the style of speech of a pre-recorded "model talker". Presumably, the same processes are at work in both cases, however, in the case of imitation, there is no personal interaction between the model talker and the speaker. There is an increasing number of recent studies investigating convergence or imitation using either (i) phonetic measures such as vowel formants [1, 2, 3, 4, 5, 6], vowel duration [1], voice onset time [7], articulation rate [5, 8], keyword duration [6], pitch [8], or spectral amplitude envelopes [9] or (ii) perceptual similarity as rated by independent listeners (e.g. [10, 11, 12]) or (iii) both phonetic measures and perceptual similarity (e.g. [7, 4, 5, 6]). However, there seem to be no consistent findings as for which phonetic features are affected in convergence. For instance, studies investigating similar research questions by way of different measures of convergence may come to different conclusions (e.g. [10] find an effect of gender on convergence using perceptual measures, while [9] using spectral amplitude envelopes finds no effect of gender). Also, studies investigating convergence using several measures may find convergence only with respect to some of these measures (e.g. [5] find perceptual convergence but not convergence with respect to articulation rate, and inconsistent behavior with respect to vowel formants). The current project aims at establishing which inventory of phonetic parameters can be affected in accommodation in general, and at investigating how consistently they are affected across speakers.

The present paper describes results obtained for one of the first parameters investigated in our project, viz. articulation rate, henceforth AR. AR has been shown to be subject to convergence already in early studies in the field [13, 14]. Interestingly, more recent studies looking at AR present results contradicting or at least relativizing these early results: as mentioned above, [5] find convergence with respect to perceptual similarity, but not with respect to AR, while [8] report convergence with respect to AR at session level, i.e. when calculating differences in rates over complete dialogs. However, effects on AR at turn level, i.e. when calculating differences turn-by-turn, fall below [8]'s required confidence threshold.

Thus the present experiment may shed more light on the role of AR in convergence. In addition, we extend these recent studies by taking social aspects into account. Our method is similar to the method first proposed by [15] for assessing convergence of turn-taking behavior. They calculated linear regression models using partner's past parameters as predictors. Similarly, [16] used linear regression in investigating not AR but other temporal parameters. He predicted participants' mutual attractiveness and competence scores. We adopt aspects of both [15] and [16]; however, we make use of linear mixed models [17] instead of traditional linear regression because they allow us to factor out random sources of variation such as speaker-specific effects. Also, we use social scores as a predictor rather than as the dependent variable.

2. Data

2.1. Data collection

We collected a corpus of 46 spontaneous conversations between female German speakers on topics of their choice. Each dialog lasted approx. 25 minutes. Participants wore head-set microphones while talking to each other in a sound-attenuated booth. We recorded the dialogs first in a unimodal (UM) and later in a multimodal (MM) condition. In the UM setting, participants could not see each other during the conversation. In the MM setting, participants could see each other through a transparent screen. We had 12 speakers in the UM condition, and 7 of them agreed to return for the MM condition. One additional speaker was recruited for the MM condition. We have 22 dialogs (approx. 10.3 hours of dialog) in the UM condition and 24 (approx. 10.5 hours) in the MM condition (cf. Table 1).

2.2. Social factors

It is well accepted that accommodation in speech is related to social factors [18, 19, 20, 16, 6]. For instance, [16] correlated the degree of convergence with speakers' mutual ratings of social attractiveness and competence. To cater for such social factors in the present database, speakers rated their conversational partner (in terms of likeability and competence) after each conversation by filling in a questionnaire. In this paper, we only

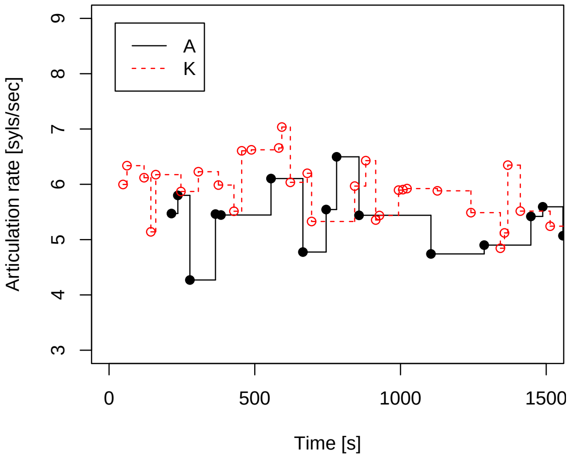

Figure 1: *AR profiles for dialog A-K from the UM condition. Rates are assumed to be constant until the next long turn.*

unimodal		unimodal		multimodal		multimodal	
dial.	#	dial.	#	dial.	#	dial.	#
A-C	5	F-E	16	A-C	17	H-D	24
A-K	23	F-J	13	A-K	12	H-F	12
B-A	8	G-B	7	A-M	16	H-J	22
B-M	18	G-L	10	C-D	17	J-A	20
C-E	18	J-D	39	C-H	11	J-F	10
C-F	12	J-I	26	C-M	20	J-K	21
D-G	12	K-C	15	D-A	12	K-C	14
D-L	18	K-F	12	D-F	2	K-H	24
E-I	31	L-B	17	D-J	20	K-M	24
E-J	26	L-M	26	F-A	14	M-D	14
		M-A	6	F-C	6	M-H	30
		M-K	36	F-K	3	M-J	29

Table 1: *Number of "long turns" per dialog. Each capital letter represents a speaker.*

assess the influence of the likeability scores. We captured mutual liking by four items in the questionnaires. Participants were asked how likeable ("sympathisch"), friendly("freundlich"), socially attractive ("sozial"), and relaxed ("locker") they found their partner on a 5-point Likert scale. We transformed the values to integers from -2 to +2. We added these values to obtain a composite score for overall likeability. Even though negative scores were rare in this experiment, this composite exhibits reasonable variation (it ranges from -2 to 8).

2.3. Parameter extraction

The recordings were automatically annotated on the segment, syllable, and word levels by forced alignment based on manually generated transliterations of the dialogs [21].

We extracted ARs turn-by-turn, defining turns as intervals of speech of a speaker with no intervening pauses of longer than 0.5 seconds. For each turn, we calculated its AR as the number of syllables divided by vocalization time, i.e. durations of turn-internal pauses were subtracted.

$$\text{AR(turn)} = \frac{\text{\#syllables(turn)}}{\text{duration(turn)-pauses}} \qquad (1)$$

The dialogs contain many very short turns, some of them backchannels such as "okay" or "u-hum", some of them short answers to very specific questions such as a German version of an exchange like "What do you study?" – "English." We wanted to avoid comparing such short turns to longer conversational turns. In shorter turns, AR would be affected to a greater extent by phrase-final lengthening. We therefore excluded all turns shorter than a certain threshold. We set this threshold to 7 seconds because this still left at least one observation for each speaker after excluding some observations for reasons described below. We will refer to the turns remaining in the analysis as "long turns" for the rest of this paper.

To assess the dynamics of the accommodation of ARs, we temporally associated each long turn's AR with the end of the turn. We then assumed that the AR remains constant until the next long turn. This reflects the idea that shorter turns do not cause the partner to re-assess her impression of the speaker's current AR. We thus receive rate profiles across dialogs for both speakers. These profiles provide the AR of each speaker's preceding long turn for any point in time in the dialog. The profiles for the dialog between speakers A and K in the MM condition are indicated as an example in Fig. 1. The dots indicate the ends of long turns; the rates remain constant until the next dot. For instance, the first long turn of speaker A ended at around 214 seconds (the first black dot at the beginning of the solid black line). At that time, K had already produced 5 long turns (the first 5 open red dots connected by the red dashed line). The AR of K's immediately preceding long turn is indicated by the red dashed line at the point where the first black dot occurs. We extracted speakers' ARs along with their partners' preceding rates for the end of every long turn in our data, i.e. in the A-K dialog for each of the dots indicated in Fig. 1.

In the example dialog, speaker A then produced her second long turn at approx. 236 seconds, before K produced another long turn (the second dot on the black solid line occurs before the next open red dot on the red dashed line, so the value for K's preceding rate is still the same (approx. 6.2). We do not want to take into account such observations because we assume that A's first long turn is more likely to be correlated with K's preceding rate than A's second long turn, as A's first long turn is intervening between the two. We therefore excluded all observations in which the partner's preceding long turn was still the same. Similarly, we excluded all observations for which we could not estimate the partner's preceding AR. In Fig. 1, this would apply to all of K's long turns which were produced before A produced her first long turn at around 214 seconds.

This procedure yielded 788 observations. Table 1 lists the number of observations for each dialog. Each extracted observation consists of a point in time in the dialog (henceforth *time*), the associated AR (*artrate*), and the partner's preceding AR (*prec.rate*). We added as factors for each observation an ID for the dialog (*dialog*), the speaker (*speaker*), the partner (*partner*), the condition (UM or MM, *modality*), and the speaker's likeability score for the partner (*liking*). In the following statistical analysis, *dialog*, *speaker*, *partner*, and *session* are treated as categorical variables, all other variables as continuous.

3. Statistical analysis

3.1. Convergence and social factors

We used R [22] and the `lme4` package [17] to analyze our data by way of linear mixed models. We follow [15] and test for convergence by using partners' preceding AR to predict a speaker's current rate: if speakers converge, the partner's pre-

Parameter	Estimate	lower	upper	p
Intercept	8.1460	7.1139	8.9882	0.0000
liking	-0.3024	-0.4373	-0.1499	0.0000
prec.rate	-0.3807	-0.5175	-0.2153	0.0000
liking:prec.rate	0.0466	0.0215	0.0690	0.0002

Table 2: *Estimated Highest Posterior Density (HPD) intervals for the fixed effects. The columns indicate the fixed effects parameters, their estimates, the lower and upper bounds for the HPD 95% intervals, and the associated p-values.*

ceding rate should be a good predictor of the speaker's current rate. Thus, speakers' AR (*artrate*) is the dependent variable, and the partners' preceding rate (*prec.rate*) is a fixed effect. As we are interested in the interference of the likeability scores, *liking* is included as a fixed effect as well. If the degree of convergence depends on *liking*, we would expect an interaction between *prec.rate* and *liking*: speakers would adapt differently to the preceding rates depending on whether they like their partner or not. As for random effects, we included *dialog*, *speaker*, and *partner* as possible random factors. The model is indicated in Equation (2). Fixed effects are indicated on the right-hand side in the first line of the equation. The random effects are indicated in the second line; for instance, (1|dialog) is the random effect notation for the factor *dialog*.

$$\begin{aligned} \text{artrate} \sim & \text{liking} + \text{prec.rate} + \text{liking} * \text{prec.rate} \\ & + (1|\text{dialog}) + (1|\text{speaker}) + (1|\text{partner}) \end{aligned} \quad (2)$$

After fitting the model, we assessed the significance of the effects by Markov Chain Monte Carlo sampling, as recommended by [23]. To this end, we used the `languageR` package [24] to estimate Highest Posterior Density (HPD) intervals and the associated p-values. The results are presented in Table 2. There is a line for each fixed effect. The columns list its estimated value ("Estimate"), the estimated HPD 95% confidence interval ("lower" is its lower bound, "upper" the upper bound), and the p-value.[1]

As indicated in Table 2, the intercept of the model corresponds to an AR of 8.1460 syllables per second. This is much higher than the average AR in all long turns (which was approx. 5.9). This is because the main effects *liking* and *prec.rate* have negative coefficients, i.e. with increased *liking* and *prec.rate*, *artrate* decreases. Therefore the intercept, which is the AR expected for a liking score of 0 and a (theoretical) preceding rate of 0, must be higher than the observed mean.

More importantly, both *liking* and *prec.rate* had highly significant effects ($p \ll 0.01$). *Liking* lowered *artrate* by 0.3024 times the liking score; as *liking* ranged between -2 and +8, this would correspond to changes in *artrate* between slightly increasing it ($-2*-0.3024 \approx 0.6$) and lowering it ($8*-0.3024 \approx -2.4$).[2] However, *prec.rate* also has a lowering effect (the coefficient was -0.3807, cf. Table 2). This is surprising: given that we are interested in convergence, one would expect a positive correlation between *prec.rate* and AR – if the preceding rate is high, convergence would require a high rate, and vice versa; this tendency should manifest itself in a positive coefficient for *prec.rate*. This is obviously not the case. However, it is not only that there was no general effect of convergence, in which case we would have obtained a coefficient of around zero, or no significant effect at all. To the contrary, the negative coefficient for *prec.rate* means that the "default" behavior of our subjects was divergence rather than convergence, i.e. in general, subjects responded to high articulation rates by lower articulation rates, and vice versa.

The most interesting outcome of fitting the model is the interaction in the last line of Table 2. The interaction between *liking* and *prec.rate* has a positive coefficient, i.e. in addition to the negative contribution of *prec.rate* discussed above, *prec.rate* also has a positive influence on *artrate*. To make the interpretation clearer, we take the model Equation (2) above, and fill in the coefficients. Since the random effects are not relevant for the interpretation, we ignore them here and obtain Equation (3). For readability we have rounded the estimated coefficients to the second decimal, and shortened the effect names: L is short for *liking*, and P stands for *prec.rate*. We then transform the term to understand the effect of *prec.rate* by factoring out P according to the distributive law in the second line:

$$\begin{aligned} \text{artrate} &\approx 8.15 - 0.30{*}L - 0.38{*}P + 0.05{*}(L{*}P) \\ &= 8.15 - 0.30{*}L + (0.05{*}L - 0.38){*}P \end{aligned} \quad (3)$$

It can then be easily seen that the negative effect of P, or *prec.rate*, may be cancelled out or reversed by high liking scores: when 0.05*L is greater than 0.38, the effect of P becomes positive. This would be the case for our highest liking score, which was 8. For lower scores, the effect of P is still negative, but what is important is that the negative effect is weakened with increasing *liking*. This shows that the influence of partner's preceding rates becomes stronger for higher liking scores, or, in other words, that there was more convergence for higher liking scores.

3.2. Dynamic effects

Several recent studies on convergence (e.g. [11, 12, 9]) assess convergence by comparing utterances early in an interaction to utterances late in the interaction. This suggests that convergence is a dynamic process which increases in the course of the interaction, and we would expect an interaction between *prec.rate* and *time*, possibly mediated by *liking*. To assess this aspect, we had first fitted a model including interactions between *time* and *prec.rate* as well as between *time* and *liking*. However, none of the interactions with *time* reached significance, and neither did *time* as a main effect. Similarly, a three-way interaction between all three effects did not improve the fit of the model, neither did taking out the unwarranted interactions but keeping *time* as a fixed effect. We will not go into the details of these first three models. It suffices to say that we could not confirm a dynamic aspect of convergence in ARs, as this should have given rise to an interaction involving *time* and *prec.rate*. All other effects observed in these models were consistent with the effects observed in Section 3.1.

[1] Since we will consider several linear mixed models below, we adopt a significance level of 0.01 in the following. However, the `languageR` package only outputs the HPD 95% interval but not the 99% interval. As we are mostly interested in the estimated coefficients and their p-values in this experiment and less interested in the exact range of the HPD interval, we still include the intervals as we consider them informative, however, readers should be aware that the intervals would be slightly greater for 99% than for 95%.

[2] This effect could be due to subjects speaking slower when feeling more at ease with a partner they like; but a more thorough investigation of possible causes is beyond the scope of this paper.

Parameter	Estimate	lower	upper	p
Intercept	7.9998	7.0501	8.9373	0.0000
liking	-0.2912	-0.4249	-0.1368	0.0001
prec.rate	-0.3970	-0.5351	-0.2288	0.0000
modality	0.5345	0.2082	0.8885	0.0023
liking:prec.rate	0.0504	0.0257	0.0736	0.0000
liking:modality	-0.0707	-0.1233	-0.0202	0.0079

Table 3: *Estimated Highest Posterior Density (HPD) intervals for the fixed effects, now including modality. The columns indicate the fixed effects parameters, their estimates, the lower and upper bounds for the HPD 95% intervals, and p-values.*

3.3. Effects of modality

As mentioned in Section 2, we had a unimodal (UM) and a multimodal (MM) condition. It has been claimed [25] that visual information enhances convergence (as measured by increased similarity in word pronunciation) in interactive tasks. Similarly, in first analyses we have found different behavior of our interactants in the UM vs. the MM condition when investigating backchannel frequency. To assess the effect of session modality, we fit another model including *modality* as a fixed factor. In this model, we did not use dialog as a random effect because its effect had been very small in the first model.[3] This model then included *modality* as a fixed main effect, as well as its interactions with *prec.rate* and with *liking*. Given [25]'s claim, we expected a positive value for the interaction of *prec.rate* with *modality*: if there is more convergence in MM dialogs, the effect of *prec.rate* should depend on *modality*. Interestingly, while the estimates for the effects of the earlier model remained relatively unchanged, neither *modality* as main effect nor its interaction with *prec.rate* reached significance. Since the two were also strongly correlated, we removed the interaction of *prec.rate* and *modality* from the model. The HPD intervals for the fixed effects of the resulting model are indicated in Table 3.

Comparing Table 3 to the previous results in Table 2, we see that the estimates and p-values for the fixed effects are still approximately the same. In the new model, *modality* has a significant effect. As *modality* was coded as 0 in UM, and as 1 in MM, this means that AR is affected by 0*0.5345 in UM, and by 1*0.5345 in the MM condition, i.e., ARs are approx. 0.5 syllables per second higher in the MM condition than in the UM condition. However, this time the interaction with *liking* exhibits a negative estimate of approx. -0.07 indicating that this effect is somewhat weakened with higher liking scores. This is in line with the negative value for *liking* as a main effect, which indicated that in general, *liking* lowered the AR. To summarize, we did not find an effect of modality on the degree of convergence, as posited by [25]. However, there was an effect of *modality* in that ARs were higher in the MM condition, but less so if speakers liked their conversation partner more.

4. Discussion & Outlook

In our perspective, convergence and divergence are not two extremes on a scale. Rather, on a continuum ranging from dissimilar speech styles to similar speech styles, we interpret a significant adjustment from left to right along this continuum as a convergence effect, and an adjustment from right to left as a divergence effect. In this paper, we have suggested that similarity of AR can be captured in the contribution of partners' preceding ARs when predicting speakers' current ARs. High similarity would entail positive coefficients for preceding ARs, and low similarity negative coefficients. An increase in the coefficients indicates convergence, a decrease indicates divergence.

Summarizing and interpreting the analyses presented above in this way, we fitted six linear mixed models. With regard to whether convergence in general occurred in our dialogs, all models yielded a negative estimate for *prec.rate* as a main effect without its interaction with *liking*, which indicates that the general tendency in the dialogs was to diverge, and not to converge. This is a surprising finding, which we would like to investigate further in the future. However, besides this negative main effect, all models also showed a highly significant positive effect of the interaction between *prec.rate* and *liking*. Thus, convergence effects on AR in this study could be observed when controlling for confounding factors such as mutual liking. This could explain the inconclusive findings of recent studies on AR ([5, 8], cf. Section 1), as they did not take into account social factors. As for the early studies on convergence in AR mentioned above, [16] investigated conversations between students and business persons of their choice, which probably entails that most interactants liked each other. The other early study mentioned above [13] used an automated interview technique: students were interviewed using prerecorded questions in which the interviewer's AR was identical across all questions in each interview. This extreme invariance in the interviewer's ARs may have given rise to more consistent effects on interviewee's ARs than can be observed in natural interactions, making the ususally subtle effects more detectable.

Concerning the dynamic aspect of convergence, we could not find a significant effect of *time* on the ARs of long turns, neither in interactions, which would have indicated that the effects of other factors change in the course of the dialog, nor as a main effect, which would have indicated that the ARs in general change in the course of the dialog. This could be due to the accommodation happening already very early in the dialogs, before the first long turns occur. In this case, more fine-grained phonetic parameters might capture the effect. Also, it is possible that the effect is not linear and can thus not be adequately captured in a linear model. We hope that future work will shed more light on this issue.

Finally, while our last two models confirm that modality does have an effect on the AR in general, it does not influence the degree of convergence, as suggested by [25]. There was no significant interaction of *modality* with *prec.rate*, which would have shown that the degree of convergence is affected by modality. However, it is interesting to note that speakers spoke faster in the MM condition, a finding that one might attribute to more lively conversations when partners were positioned face-to-face. Future work investigating other prosodic parameters such as pitch range or pitch accent frequency may help to answer this question.

5. Acknowledgements

This study is part of the project *Phonetic Convergence in Spontaneous Speech* within the SFB 732 funded by the German Research Foundation (DFG).

[3] Indeed, refitting the previous model without dialog as a random factor did not change the general outcome, it just yielded slightly different estimates for the fixed effects, but no difference in the significance of these effects.

6. References

[1] V. Delvaux and A. Soquet, “The influence of ambient speech on adult speech productions through unintentional imitation,” *Phonetica*, pp. 145–173, 2007.

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

[3] M. Babel, “Dialect divergence and convergence in New Zealand English,” *Language in Society*, vol. 39, no. 4, pp. 437–456, 2010.

[4] J. S. Pardo, “Expressing oneself in conversational interaction,” in *Expressing oneself/expressing one’s self: Communication, cognition, language, and identity*, E. Morsella, Ed. London: Taylor & Francis, 2010, pp. 183–196.

[5] J. S. Pardo, I. Cajori Jay, and R. M. Krauss, “Conversational role influences speech imitation,” *Attention, Perception, & Psychophysics*, vol. 72, no. 8, pp. 2254–2264, 2010.

[6] 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.

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

[8] 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.

[9] N. Lewandowski, *Talent in nonnative phonetic convergence*. Doctoral dissertation, Universität Stuttgart, 2011. [Online]. Available: http://elib.uni-stuttgart.de/opus/volltexte/2012/7402/pdf/Lewandowski.pdf

[10] L. Namy, L. Nygaard, and D. Sauerteig, “Gender differences in vocal accommodation: The role of perception,” *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, 2002.

[11] J. S. Pardo, “On phonetic convergence during conversational interaction,” *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.

[12] M. Kim, W. S. Horton, and A. R. Bradlow, “Phonetic convergence in spontaneous conversations as a function of interlocutor language distance,” *Journal of Laboratory Phonology*, vol. 2, pp. 125–156, 2011.

[13] 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.

[14] 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.

[15] J. N. Cappella and S. Planalp, “Talk and silence sequences in informal conversations III: Interspeaker influence,” *Human Communication Research*, vol. 7, no. 2, pp. 117–132, 1981.

[16] R. Street, “Speech convergence and speech evaluation in fact-finding interviews,” *Human Communication Research*, vol. 11, no. 2, pp. 139–169, 1984.

[17] D. Bates, M. Maechler, and B. Bolker, *lme4: Linear mixed-effects models using S4 classes*, 2012, R package version 0.999999-0. [Online]. Available: http://CRAN.R-project.org/package=lme4

[18] 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.

[19] 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.

[20] 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.

[21] S. Rapp, “Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models—An aligner for German,” in *Proceedings of ELSNET Goes East and IMACS Workshop ”Integration of Language and Speech in Academia and Industry” (Moscow, Russia)*, 1995.

[22] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[23] R. Baayen, *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge University Press, 2008.

[24] R. H. Baayen, *languageR: Data sets and functions with ”Analyzing Linguistic Data: A practical introduction to statistics”*, 2011, R package version 1.4. [Online]. Available: http://CRAN.R-project.org/package=languageR

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