Exemplar Dynamics in Phonetic Convergence of Speech Rate

Antje Schweitzer, Michael Walsh

Institute for Natural Language Processing
University of Stuttgart, Germany
{antje.schweitzer,michael.walsh}@ims.uni-stuttgart.de

Abstract

We motivate and test an exemplar-theoretic view of phonetic convergence, in which convergence effects arise because exemplars just perceived in a conversation are stored in a speaker's memory, and used subsequently in speech production. Most exemplar models assume that production targets are established using stored exemplars, taking into account their frequency- and recency-influenced level of activation. Thus, convergence effects are expected to arise because the exemplars just perceived from a partner have a comparably high activation. However, in the case of frequent exemplars, this effect should be countered by the high frequency of already stored, older exemplars. We test this assumption by examining speech rate convergence in spontaneous speech by female German speakers. We fit two linear mixed models, calculating speech rate on the basis of either infrequent, or frequent, syllables, and predict a speaker's speech rate in a phrase by the partner's speech rate in the preceding phrase. As anticipated, we find a significant main effect indicating convergence only for the infrequent syllables. We also find an unexpected significant interaction of the partner's speech rate and the speaker's assessment of the partner in terms of likeability, indicating divergence, but again, only for the infrequent case.

Index Terms: phonetic convergence, speech rate, exemplar theory, spontaneous speech, frequency effects

1. Introduction

Phonetic convergence is the process of adapting one's speech to an interlocutor. The opposite, i.e., the adoption of a speech style that differs from that of the interlocutor, is called divergence. In both cases, the perception of the interlocutor's speech affects a speaker's current production targets. According to Communication Accommodation Theory (CAT, e.g. [1, 2, 3, 4]), the adaptation seen in convergence or divergence is a dynamic process and affects not only speech, but communicative behavior in general (i.e. linguistic and phonetic features, but also paralinguistic aspects). CAT proposes that convergence decreases social distance between interlocutors and thus reflects a speaker's (often unconscious) need for social integration or identification with the interlocutor's social group [2]. In contrast, divergence is caused by the need to distance oneself from the interlocutor's group. Interlocutors may also converge to increase intelligibility and efficiency of communication [5, 6, 7]. Thus, social factors and the communication setting are clearly important when investigating convergence.

As a complementary approach, we suggest an exemplar-theoretic explanation in which convergence arises as a consequence of the speech production/perception loop and the role of stored exemplars in this process. Exemplar-theoretic accounts of speech perception/production [8, 9, 10, 11] typically contend that listeners categorise and store perceived speech exemplars in memory. This gives rise to richly detailed clouds of exemplars in the mind of the listener, from which production targets are derived in subsequent speech production. Exemplar categories whose tokens occur frequently will have densely populated clouds, whereas categories that are rarely employed will have few exemplars stored. This entails sensitivity to frequency both in speech production and perception, and indeed exemplar models of speech have successfully accounted for frequency effects across multiple linguistic levels [8, 9, 12, 13, 11]. However, to our knowledge no research has examined the possibility that convergence might be sensitive to frequency.

We address this by examining convergence of speech rate in a recently developed corpus of spontaneous German speech [14, 15]. Our hypothesis is that convergence effects are more likely to be detectable for infrequent than for frequent exemplars, because a recently perceived exemplar stored in a sparse cloud is likely to be proportionately more influential in subsequent productions than an exemplar stored in a dense cloud.

Given the previous work in CAT concerning the role of social factors, we also examine the relationship between how dialog partners view each other, in terms of likeability and competence, and how this view might influence speech rate behaviour and interact with exemplar frequency.

We proceed as follows: In section 2 we address related work on convergence of speech rate, and in section 3 we describe the speech data in more detail. We then outline our methodology and statistical analysis for assessing convergence in speech rate in section 4. In section 5 we present our results, and in section 6 we provide a detailed account of how exemplar dynamics can account for our findings.

2. Related Work

Speech rate has been shown to be subject to convergence in some of the first studies on communication accommodation [16, 17]. Interestingly, more recent studies looking at speech rate present results contradicting or at least relativising these early results: [18] find convergence with respect to perceptual similarity, but not with respect to speech rate, while [19] report convergence in a dialog task with respect to speech rate at session level, i.e. when calculating differences in rates over complete dialogs. However, when calculating differences turn-by-turn, the effects fall below their required confidence threshold. Finally, [14] found an unexpected general effect of divergence, rather than convergence, when looking at speech rate turn-by-turn; however, this was countered by a convergence effect in the case where participants liked each other.

Given these diverse findings, the present study aims to investigate convergence of speech rate further, this time taking frequency of occurrence into account as a potential influencing

factor. In the current study, we adopt a method employed by [20, 14], and test for convergence (or divergence, respectively) by using a partner's preceding speech rate to predict a speaker's current rate: if speakers converge (or diverge), the partner's preceding rate should be a good predictor of the speaker's current rate. I.e. in both cases the adaptation is indicated by a significant effect of the partner's preceding rate on the speaker's current rate. In the case of convergence, the direction of this effect is positive, and in the case of divergence, the effect is negative. [14] focused on longer conversational turns, reasoning that they are less affected by phrase-final lenthening. They also sought to avoid comparing longer conversational turns to very short turns such as backchannels. However, this meant (i) they had relatively few data per speaker, and (ii) the turn from which the preceding rate was taken as a predictor, may have occurred much earlier than the current turn, potentially minutes earlier. Our approach differs, however, in that we are less restrictive with respect to the required minimum length of turns, and our turns are always temporally contiguous, thus taking recency into account. We discuss the specific details in section 4.

3. The GECO speech database

In the current study, we investigate speech rate in the GECO [15] database. GECO consists of spontaneous conversations between previously unacquainted female German speakers on topics of their choice. Each dialog lasted approx. 25 minutes. Participants wore headsets while talking to each other in a sound-attenuated booth. There are 46 dialogs (approx. 21 hours of dialog) in the database. The recordings were automatically annotated on the segment, syllable, word, and prosodic levels. The resulting corpus amounts to approx. 250,000 words, 360,000 syllables, and 870,000 phones.

It is well accepted that accommodation in speech is related to social factors [1, 2, 3, 21, 22]. For instance, [21] correlated the degree of convergence with speakers' mutual ratings of social attractiveness and competence. The GECO database provides similar ratings: after each conversation speakers rated their conversational partner, e.g. in terms of likeability and competence, by filling in a questionnaire. Likeability and competence were captured by four items each, with a 5-point Likert scale for each item.[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 and those of the four competence items to obtain composite scores for each of the two aspects. Even though negative scores are rare, the composite scores exhibit reasonable variation (both range from -2 to 8).

4. Method

In order to assess how one speaker's speech rate influences an other speaker's speech rate, we used linear mixed models (the *lme4* package [23] in R [24]) to predict speech rates in each turn-initial phrase of a speaker by the immediately preceding turn-final phrase of the other speaker.[2] Since we expect that convergence effects should be stronger for infrequent than for frequent syllables, we performed this analysis twice, once taking only infrequent syllables into account for calculating the speech rate, and once taking only frequent syllables into account. Additional factors that we expected to impact upon speech rate were the original number of syllables in the phrase, and the likeability and competence scores that speakers and partners assigned to each other.

4.1. Frequent and infrequent syllables

Our first step was to establish frequency counts for every syllable type. Given the considerable size of the corpus, we believe these frequency counts are reasonably representative of what one might find in conversational speech in general.

To classify syllables as frequent or infrequent, we partitioned them into quartiles based on their frequency counts. In the lowest quartile, we have syllables with frequencies ranging from 1 to 322, corresponding to 3,944 types. We consider these the infrequent syllables in the database, while we took the upper three quartiles, i.e. 75% of the data, as frequent syllables.[3] These upper three quartiles then contained 202 syllable types, with frequencies between 324 and 10,946.

4.2. Calculating speech rate

We aim to establish whether the speech rate of turn-initial phrases by speaker A is influenced by the speech rate of the preceding turn-final phrase by speaker B and the role of frequency in this relationship. For the purpose of statistical modeling, we created two sets of data; a *frequent* set, where we estimated the speech rate for each phrase taking only frequent syllables into account, and an *infrequent* set, using only infrequent syllables. Phrases containing frequent and infrequent syllables were represented with separate speech rate values in both sets, one based on the frequent syllables, the other on the infrequent syllables.

For both sets, for each phrase we divided the number of syllables taken into account by their total duration. For instance, in the infrequent set we calculated speech rate for a phrase (potentially composed of both infrequent and frequent syllables) using only the infrequent syllables in that phrase, by dividing their number by their total duration, and ignoring frequent syllables. In order to avoid effects of phrase-final lengthening, syllables before pauses were excluded from all calculations. We also removed phrases that contained hesitations, laughter, or implausible durations that could indicate alignment errors.

4.3. Statistical analysis

Before fitting the linear mixed models, we removed outliers in both sets independently, on a speaker-by-speaker basis, with reference to the interquartile range: we removed all data points where either the turn-initial speech rate (the dependent variable) or the turn-final speech rate (the predictor variable) was more than 1.5 times the interquartile range below the first quartile or above the third quartile. We also removed data points where either of the two speech rate variables had been calculated on the basis of one single syllable.[4] In addition we centered all vari-

[1] See [14, 15] for details concerning the items.

[2] Note that it is not possible to identify the moment in the course of the dialog when convergence begins; instead it makes the simplifying assumption that the size of the convergence effect remains constant throughout the dialog.

[3] In the top quartile, there were only 16 syllable types. This reflects the Zipfian properties of the data, with large numbers of rare syllables and extremely small numbers of very frequent syllables. Given the very low number of syllable types in the upper quartile, many of these syllable instances might come from a very small number of words. In order to avoid lexical effects, we decided to not take just the top quartile as frequent syllables. However, analysis using just syllables from the upper quartile yielded the same findings.

[4] Speech rates on the basis of a single syllable are particularly prone to problematic effects such as the influence of syllable complexity, and of alignment errors. These errors occur less often on longer stretches of speech, as greater contextual information improves alignment perfor-

ables. The final data sets comprised 2,622 data points for the *infrequent* set, and 6,259 data points for the *frequent* set.

On both data sets, we tested for convergence by fitting a linear mixed model in which the dependent variable was the speech rate of a turn-initial phrase, and the partner's speech rate from the immediately preceding turn-final phrase was a predictor variable. As previous studies on convergence have shown that social factors play a role, we also included the mutual likeability and competence ratings as predictors. We expected that the impact of the partner's speech rate will be stronger when the speaker likes the partner more, or thinks the partner is more competent. Thus, in both models we needed pair-wise interactions between the speech rate of the preceding phrase and the likeability and competence scores. A further factor, though one that is not in the focus of our investigation here, is the original number of syllables in the phrase. We included it (counting both infrequent and frequent syllables as well as final syllables) as we expected it to improve the fit of the models by explaining some of the variation in speech rates. As random effects, we specified intercepts for speaker. This amounts to assuming that individual speakers have individual speech rates, i.e. some speakers are expected to just speak slower or faster in general. In addition, we specified by-speaker slopes for the likeability and competence scores. The slopes allow individual speaker "habits" of reacting to more likeable, or more competent, partners by speaking slower in general, or faster in general.

The formula for the model is given in (1) below.

$$\begin{aligned} rate \sim\ & number\ of\ original\ syllables \\ & + preceding\ rate \\ & + preceding\ rate : likeability \\ & + preceding\ rate : competence \\ & + (1 + competence + likeability \mid speaker) \end{aligned} \tag{1}$$

We fit this model once on the infrequent data set, and once on the frequent data set. Visual inspection of the residual plots revealed no heteroscedasticity for either data set.

Our hypothesis is that the preceding rate and its interactions are only significant predictors on the infrequent data set. Since we run the analysis on two sets of data, and will assess significance for four factors in each set, we apply Bonferroni correction for 8 tests, which gives us $\alpha = 0.05/8 = 0.00625$. In order to provide p-values for the factors in our models, we used likelihood ratio tests in the following way. For each factor, we fit a model that differed from the full model only in that the factor in question was removed. We then compared the two models by pairwise anovas. A factor was considered significant if (i) the p-value provided by the anova was below α, and if at the same time (ii) the AIC value of the full model was at least 2 points smaller than the simpler model without the factor. The fact that the full model provides a significantly better fit was taken to indicate that the factor is a significant predictor.

5. Results

Table 1 provides coefficients, t-values, and p-values for the models on the infrequent (top panel) and on the frequent (bottom panel) data set. In the infrequent case all factors, with the exception of the interaction between preceding rate and competence, were significant. As anticipated, the number of original syllables in the current phrase was a signficiant predictor of the speech rate for the phrase. Our main interest, however, is the effect of the preceding rate. As indicated in the table, it significantly affects the current rate. The coefficient of 0.043 indicates that the effect size is small (see below). This subtle effect is in keeping with what was found for other parameters in previous work [15]. The fact that the coefficient is positive shows that we observe convergence rather than divergence: The higher the speech rate in the preceding phrase, the higher the speech rate in the current phrase. Interestingly, the interaction between preceding rate and likeability yields a negative coefficient, i.e., divergence, thus running counter to the influence of preceding rate alone. No significant effect was found for the interaction between preceding rate and competence.

Table 1: *Fixed factors, their estimated coefficients, t-values, and significance at a level of $\alpha = 0.00625$ in the two final models. The effects of preceding rate and its interaction with likeability are only significant on the infrequent data set, confirming our hypothesis that infrequent syllables should exhibit convergence effects more easily.*

Infrequent data set			
Factor	Coeff.	t-value	Sign.
number of orig. syllables	0.007	2.631	*
preceding rate	0.043	2.352	*
preceding rate : likeability	-0.027	-2.906	*
preceding rate : competence	0.020	1.959	n.s.
Frequent data set			
Factor	Coeff.	t-value	Sign.
number of orig. syllables	0.051	17.871	*
preceding rate	0.017	1.447	n.s.
preceding rate : likeability	-0.008	-1.347	n.s.
preceding rate : competence	0.012	1.914	n.s.

Unlike in the infrequent data set, neither of the factors that would indicate convergence (or divergence), i.e. neither the preceding rate nor its interaction with likeability and competence, significantly affect the speech rate in the frequent data set.[5]

6. Discussion

As indicated in section 1, convergence or divergence effects can be explained by exemplar dynamics. According to exemplar models [8, 9, 11], perceived exemplars are stored in memory with a great amount of phonetic detail. Instances which are similar to each other, across particular dimensions (e.g. formant characteristics, intensity, duration etc.), will lie closer together in the perceptual space spanned by those dimensions than those which are less similar. Thus exemplars of the same category will be close to each other; they will form clouds of similar exemplars. It is usually assumed that each exemplar is activated to some degree, and that the level of activation depends on how recently the exemplar has been stored and/or accessed. Exemplars are subject to decay when their activation falls below some threshold. Hence the clouds vary constantly as new exemplars are stored, and old unused exemplars fade from memory.

Exemplar models assume that in speech perception newly perceived tokens are categorised on the basis of the stored exemplars and their categories: similarity of an incoming token to existing exemplars activates these stored exemplars, and the

mance. Calculating speech rates on the basis of more syllables provides a smoothing effect to counter these problems.

[5] A reviewer queried whether there was an overall effect of preceding rate. We did not find any significant effect of preceding rate when frequency was ignored.

new token is then categorized accordingly. Some models, for instance [25, 8], consider activation of a category as a cumulative function over the category's members. Thus, incoming exemplars would activate several categories, to varying degrees, by activating some of their members, and the new token would then be categorised and stored as belonging to the category with the highest cumulative activation.

Similarly, in speech production, exemplars of the target category are activated, and a production target is formed, either by averaging (possibly weighted) over the activated exemplars, or by selecting one of the activated exemplars at random [9, 11]. The production process is at the core of an exemplar-theoretic explanation of convergence: Recently perceived exemplars, i.e. exemplars just perceived when listening to the partner, are still highly activated, and thus automatically contribute to a speaker's next production targets. However, the specific contribution of such a recent exemplar depends on how many other exemplars also contribute in production. Frequent categories are densely populated because they are regularly updated when an individual encounters and stores new tokens of the category, or produces a token of the category. On the other hand, infrequent categories have sparse exemplar clouds because their tokens are rarely availed of for production or categorisation, and hence decay, and the individual rarely encounters new instances of the category. Hence, the influence a token has on the activation of the category as a whole is proportionately greater if the category is infrequent, i.e. has a sparsely populated cloud, than if it is frequent.

Given this account of convergence in exemplar-theoretic terms, the case of our speech rates for the German data examined above can be explained as follows. During the speaker's perception of the partner's previous phrase, the perceived syllables become members of the speaker's syllable clouds. They subsequently affect the speaker's selection of syllable exemplars in production. When producing a syllable of the same type as a recently perceived syllable, the speaker selects a production target either by random from, or by averaging over, the exemplars in a syllable exemplar cloud in which a syllable from the previous phrase has just been stored. The new member of the category is likely to be far more influential if the category is sparsely populated. Consequently we only find this convergence effect in the infrequent data set.

It is also possible that the syllables in the previous phrase and the current phrase are not of the same type, and, hence, the syllable clouds accessed for the current phrase will not contain a recent exemplar. However, exemplar categorisation occurs across multiple dimensions and activates exemplars of several categories. As a result, the durational characteristics of a syllable categorised during the previous phrase should activate syllables with similar durational properties when producing the current phrase, even if those categories are not of the same type. That is, although the percepts from the previous phrase do not match the productions in the current phrase along the dimensions characterising syllable type, they match along the dimension of duration and thus affect exemplar selection on that dimension. For instance, perceiving a short syllable would raise the activation level of other short syllables even though they are not of the same type. Again, this is likely to have a larger impact if the density of the exemplar cloud being accessed is low, which is the case only for the infrequent data set.

Unlike our main effect, the significant interaction between preceding rate and likeability yielded by our statistical modelling is an unexpected finding, as it indicates that speech partners diverge if they like each other. The effect size is slightly larger than that of the main effect of convergence: The coefficient for the main effect of preceding rate was 0.043 (cf. table 1). Since preceding rates range from approx. -3.4 to 4.8 after centering, the main effect of preceding rate can be quantified as lying in the range from -3.4 * 0.043 (for the slowest preceding rate) to 4.8 * 0.043 (for the highest preceding rate), i.e. between -0.15 and 0.21, indicating that a speaker's rate is reduced/increased within the range of -0.15 to 0.21 syllables per second, depending on the preceding rate. The coefficient for the interaction between preceding rate and likeability was -0.027. Likeability ranged from -7.4 to 2.6 (after centering), which means that the interaction contributes a value between -7.4 * -0.027 times the preceding rate (for least likeable partners) and 2.6 * -0.027 times the preceding rate (for most likeable partners). To give an example, the most frequent likeability score in our data was approx. 2.5. In this case, the effect of the interaction lies between 2.5 * -0.027 * -3.4 = 0.23 (for the slowest preceding rate) to 2.5 * -0.027 * 4.8 = -0.32 (for the highest preceding rate), i.e. an adjustment between 0.23 and -0.32 syllables per second. It is important to note that this is a divergence effect: it increases the current rate in case of low preceding rates and decreases the current rate in case of high preceding rates.

It is unclear what causes this unexpected divergence effect when likeability is included. Interestingly, it is again only present in the infrequent case, confirming our hypothesis that convergence is subject to frequency effects. It should be noted that phonetic convergence has been claimed to be a means to reduce the social gap between partners [2]. However, in the case of speech partners who like each other, one could argue that convergence is unnecessary as there might be no social gap to bridge. In this case, their level of comfort in their partner's company would allow them to assert their own personality more assuredly. This would obviate the need for convergence, but it does not necessarily license divergence. In any case, this needs to be examined further in future work.

Regarding the absence of a significant interaction between competence and preceding rate, a previous study [21] found that convergence of speech rate was related to higher competence ratings, but not to higher social attractiveness ratings, in a scenario where students interviewed business persons and professionals. We however find the opposite pattern, perhaps because the speakers in the GECO database were social peers in a more social scenario, in which likeability is more important, whereas the participants in [21] were professionals in a professional scenario, where competence plays a greater role. Clearly, like the impact of likeability, this needs to be investigated further.

7. Conclusion

At the outset we hypothesised that phonetic convergence, like other phenomena in phonetics and phenomena found in other linguistic domains, might be subject to effects of frequency. The results of our statisical modelling, interpreted from an exemplar-theoretic perspective, confirm that for speech rate convergence this is indeed the case. Furthermore, our analysis revealed a complex relationship between social scores and exemplar dynamics which we will examine in the future.

8. Acknowledgements

This study is part of the projects *Phonetic Convergence in Spontaneous Speech* and *Exemplar-based Speech Representation* within the SFB 732 funded by the German Research Foundation (DFG).

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] K. Johnson, "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, K. Johnson and J. W. Mullennix, Eds. San Diego: Academic Press, 1997, pp. 145–165.

[9] J. Pierrehumbert, "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency and the Emergence of Linguistic Structure*, J. Bybee and P. Hopper, Eds. Amsterdam: Benjamins, 2001, pp. 137–157.

[10] ——, "Probabilistic phonology: Discrimation and robustness," in *Probability Theory in Linguistics*, R. Bod, J. Hay, and S. Jannedy, Eds. The MIT Press, 2003, pp. 177–228.

[11] M. Walsh, B. Möbius, T. Wade, and H. Schütze, "Multilevel exemplar theory," *Cognitive Science*, vol. 34, pp. 537–582, 2010.

[12] J. Bybee, "From usage to grammar: The minds response to repetition," *Language*, vol. 84, pp. 529–551, 2006.

[13] K. Schweitzer, M. Walsh, S. Calhoun, H. Schütze, B. Möbius, A. Schweitzer, and G. Dogil, "Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation," *Speech Communication*, vol. 6, pp. 65–81, Feb. 2015.

[14] 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.

[15] ——, "Social factors in convergence of F1 and F2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production, Cologne*, 2014.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[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] R. M. Nosofsky, "Attention, similarity, and the identification-categorization relationship," *Journal of Experimental Psychology: General*, vol. 115, no. 1, pp. 39–57, 1986.