Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments

Conference Paper · September 2019

DOI: 10.21437/Interspeech.2019-2445

5 authors, including:

Iona Gessinger
Universität des Saarlandes
15 PUBLICATIONS **58** CITATIONS
SEE PROFILE

Bistra Andreeva
Universität des Saarlandes
87 PUBLICATIONS **552** CITATIONS
SEE PROFILE

INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments

Iona Gessinger[1], *Bernd Möbius*[1], *Bistra Andreeva*[1], *Eran Raveh*[1], *Ingmar Steiner*[2]

[1]Language Science and Technology, Saarland University, Germany
[2]audEERING GmbH, Gilching, Germany

gessinger@coli.uni-saarland.de

Abstract

This paper discusses phonetic accommodation of 20 native German speakers interacting with the simulated spoken dialogue system Mirabella in a Wizard-of-Oz experiment. The study examines intonation of wh-questions and pronunciation of allophonic contrasts in German. In a question-and-answer exchange with the system, the users produce predominantly falling intonation patterns for wh-questions when the system does so as well. The number of rising patterns on the part of the users increases significantly when Mirabella produces questions with rising intonation. In a map task, Mirabella provides information about hidden items while producing variants of two allophonic contrasts which are dispreferred by the users. For the [ɪç] vs. [ɪk] contrast in the suffix ⟨-ig⟩, the number of dispreferred variants on the part of the users increases significantly during the map task. For the [ɛː] vs. [eː] contrast as a realization of stressed ⟨-ä-⟩, such a convergence effect is not found on the group level, yet still occurs for some individual users. Almost every user converges to the system to a substantial degree for a subset of the examined features, but we also find maintenance of preferred variants and even occasional divergence. This individual variation is in line with previous findings in accommodation research.

Index Terms: human-computer interaction, phonetic accommodation, Wizard-of-Oz experiment

1. Introduction

The interaction with computers via spoken language is becoming more and more present in our everyday life. In this context, the question arises whether humans behave similarly when speaking to a fellow human or a computer. With phonetic accommodation this paper is targeting a phenomenon that is well-documented for human-human interaction [1, 2, 3, 4] and assumed to have a strong social component: by converging to, or diverging from, our interlocutor, we are communicating a closer or more distant relationship, respectively [5]. As such one might intuitively expect that it does not necessarily occur in human-computer interaction (HCI). However, the status of computers as social actors was found to be relevant early on [6] and is discussed in the context of linguistic accommodation [7]. With speech synthesis development striving for more naturalness and interactions with spoken dialogue systems (SDSs) moving from simple commands towards free conversations, it can only be assumed that this status will further establish itself.

Thinking in terms of advantages for HCI, phonetic convergence on the part of the human towards the speech output of the computer, hence becoming more similar to it, would only be beneficial if this speech output corresponds to the speech input expected by the automatic speech recognition (ASR). For example, the optimal speaking rate for the ASR to process an incoming speech signal could be the same as the speaking rate of the computer's speech output.

The user on the other hand, may benefit from phonetic convergence towards the computer in a learning context, e.g., in computer-assisted language learning (CALL). Here, especially the pronunciation of speech segments and the realization of prosodic phenomena such as question intonation, lend themselves as targets for convergence that would lead to an improvement in the production of the learned language. We will examine such features in the present paper.

Developing computers who are themselves able to phonetically accommodate to the user is complementary to the research presented here. Specifically for the application in CALL, a synergy of the computer recognizing erroneous productions of the user, diverging from them to give room for accommodation and, eventually, the user converging to the computer, would probably be an ideal solution.

In the present study, we apply the Wizard-of-Oz (WOz) method to simulate an intelligent SDS. While the user believes to interact with an autonomous system, it is in fact the *wizard*, i.e., the experimenter, who makes decisions about the system's responses. This method has been previously used in the context of phonetic accommodation at the suprasegmental level [8, 9, 10]. However, to the best of our knowledge, this is the first study examining accommodation to question intonation and segment realization with the WOz method.

2. Method

We examine the data of 20 native German speakers (aged 18 to 55 years, mean age 27 years, 16 female) who took part in a WOz experiment during which they believed to interact with the intelligent SDS *Mirabella*. However, Mirabella's utterances were actually pre-recorded by a female native German speaker (aged 26 years) and manually played back to the users of the system by the experimenter. Mirabella is introduced as a tutor for German as a foreign language. Note that the native German speaking participants are instructed to test the system for later use with learners of German.

In this paper, we present the individual tasks as they pertain to the two phenomena under examination: intonation of wh-questions and pronunciation of the German allophonic contrasts [ɪç] vs. [ɪk] as a realization of the suffix ⟨-ig⟩ and [ɛː] vs. [eː] as a realization of long, stressed ⟨-ä-⟩. Both phenomena have two tasks pertaining to them, with the first determining the user's baseline preference and the second testing for accommodation towards Mirabella. Task numbering refers to order of presentation during the WOz experiment; for a full chronological overview of the interaction with Mirabella, refer to Gessinger *et al.* [11].

http://dx.doi.org/10.21437/Interspeech.2019-2445

2.1. Intonation

Baseline (task 2) The user formulates five constituent wh-questions whose components are given as fragments, e.g., *was – die Hauptstadt – von Lettland – sein* (what – the capital – of Latvia – be). The questions are answered by Mirabella.

Testing (task 3) Mirabella and the user take turns asking and answering each other about ten animals hiding in ten houses on the screen in the following form:

Q: *Wo hat sich* ⟨***the animal***⟩ *versteckt?*
Where did ⟨the animal⟩ hide?

A: ⟨***the animal***⟩ *hat sich in Haus Nummer* ⟨***no.***⟩ *versteckt.*
⟨the animal⟩ hid in house number ⟨number⟩.

The task is divided into two rounds of 20 turns. Mirabella produces all questions with a nuclear pitch accent on the respective ⟨*animal*⟩ followed by final *f0* fall in round 1, and a nuclear pitch accent on the interrogative pronoun *wo* followed by final high *f0* rise in round 2. The order in which Mirabella and the user ask for the animals on the screen is free.

Prediction The general unmarked expectation is for German wh-questions to be produced with falling intonation [12]. Rising intonation is mainly expected in the case of echo questions [13], i.e., when the answer was not understood and the question is uttered again. In the context of the question-and-answer exchange at hand, such echo questions are unlikely to occur, since the answers do not necessarily have to be understood by the user: the correct pictures are always visually marked on the screen as well. In the second round of the exchange, the animals remain paired with the same house numbers as before, but the pictures are arranged in a different order on the screen. Therefore, it is unexpected, yet not pragmatically wrong, to ask for the location of the animals with rising intonation. We expect to find falling intonation contours for the questions in task 2 and the first round of task 3, and a substantial increase of rising contours from the first to the second round of task 3.

Evaluation The intonation contours of 526 questions uttered in task 2 and task 3 were perceptually classified by two trained phoneticians as falling, rising, or "other contour" (depending on occurrence) taking the position of the nuclear pitch accent into account.

2.2. Segments

Baseline (task 1) The user names pictures and translates English adjectives to German by uttering them in the carrier sentence: *Das Wort* ⟨***item***⟩ *kenne ich.* (The word ⟨item⟩ is known to me.) These items contain the target words for the [ɪç] vs. [ɪk] and [ɛː] vs. [eː] contrasts. The individual realizations are perceptually categorized by the experimenter to determine the user's preferences.

Testing (task 4) The user describes all objects on a map from leaving a house until reaching a destination as follows:

a) *Ich gehe um die Maus herum. Die Maus ist* ***lustig****.*
I am walking around the mouse. The mouse is **funny**.
► bold target contains [ɪç] vs. [ɪk] contrast

b) *Ich gehe an dem* ***Käse*** *vorbei. Der* ***Käse*** *ist alt.*
I am walking past the **cheese**. The **cheese** is old.
► bold target contains [ɛː] vs. [eː] contrast

Some of the pertinent items are hidden behind boxes. The user asks Mirabella about these items and she provides the missing information while using the users's dispreferred variant of the respective pronunciation contrast (as determined by the experimenter in task 1). Given this information, the user can formulate the required utterance. If the target item is an object, it will occur twice in the utterance (*Käse* in the example above); if the target item is an adjective, it will occur only once, in the second part of the utterance (*lustig* in the example above). The task consists of four maps and contains a total of 12 occurrences per allophonic contrast.

Prediction The two allophonic contrasts examined here are regionally distributed with [ɪç]–[eː] being predominant in the North and [ɪk]–[ɛː] in the South of the German-speaking region. However, the contrasts are not believed to be strong dialectal markers. [ɪç]–[ɛː] are furthermore codified Standard German. Often, the opposite is thought to be the case by speakers, since the written form of the suffix ⟨-ig⟩ hints towards [ɪk] being the standard and there is a tendency of long, stressed ⟨-ä-⟩ merging to [eː] across the German-speaking region. It can be assumed that there is more awareness about the regionality of the [ɪç] vs. [ɪk] contrast.

Note that all native German speakers should be able to produce both variants of both contrasts, since we consider fricative variants such as [ʃ] or [ɕ] as part of the potentially problematic [ɪç] category. Therefore, all users have the means to converge to Mirabella.

We expect to find a substantial increase of the dispreferred variant for the [ɪç] vs. [ɪk] contrast and a substantial shift in the F1-F2 space in the direction of the dispreferred variant for the [ɛː] vs. [eː] contrast during the map task as compared to the baseline task.

Evaluation The realizations of the suffix ⟨-ig⟩ uttered in task 1 and task 4 were perceptually classified as belonging to the fricative or plosive category by a trained phonetician. Since speakers are not always consistent in using only one variant during the baseline task, preference reflects the majority variant produced during task 1. Individual realizations were further classified as being the same as or a different variant than the one produced by Mirabella.

For all realizations of long, stressed ⟨-ä-⟩ uttered by the users in task 1 and task 4 as well as by Mirabella, the first and second formants were measured at midpoint using Praat's [14] Burg algorithm. In a second step, Euclidean distance (*dist*) in the F1-F2 space between each user realization (*U*) and the respective realization by Mirabella (*M*) was calculated (Equation (1) for baseline task, Equation (2) for map task). Finally, difference in Euclidean distance (*dDist*) between baseline task and map task was calculated (Equation (3)).

$$dist(b) = \sqrt{(U_{baseF1} - M_{F1})^2 + (U_{baseF2} - M_{F2})^2} \quad (1)$$

$$dist(m) = \sqrt{(U_{mapF1} - M_{F1})^2 + (U_{mapF2} - M_{F2})^2} \quad (2)$$

$$dDist = dist(b) - dist(m) \quad (3)$$

Difference in Euclidean distance has the following potential outcomes:

$dDist = 0$ if the users do not shift their productions in the F1-F2 space (maintenance);

$dDist > 0$ if the users shift their productions in the direction of Mirabella (convergence);

$dDist < 0$ if the users shift their productions away from Mirabella (divergence).

3. Results

3.1. Intonation

Figure 1 shows the results of the perceptual evaluation of the question intonation contours. Three contour types were found in the data: **falling**, **falling-rising**, and **rising** [13]. The lat-

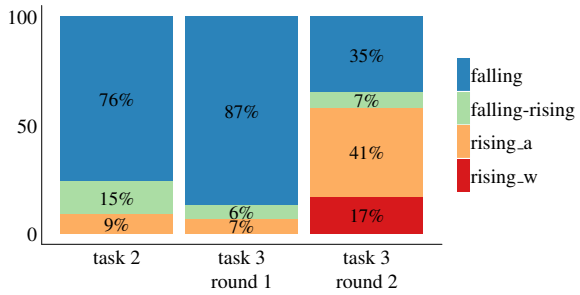

Figure 1: *Occurrence of the three observed intonation patterns during tasks 2 and 3. The rising pattern occurs with nuclear pitch accent on the animal (**rising_a**) or the interrogative pronoun (**rising_w**). In round 1 of task 3 Mirabella produces **falling** questions, in round 2 **rising_w** questions.*

ter occurs in two variants: with nuclear pitch accent on the respective *animal* (**rising_a**) or on the interrogative pronoun *wo* (**rising_w**).

In task 2, where wh-questions were formulated from fragments, **falling** contours are predominant with 76 % of all instances, but **falling-rising** (15 %) and **rising_a** (9 %) contours are produced as well. In the first round of task 3, where Mirabella produces **falling** contours, the predominance of **falling** contours on the part of the users becomes more pronounced (87 %), yet **falling-rising** (6 %) and **rising_a** (7 %) contours still occur. In the second round of task 3, where Mirabella produces **rising_w** contours, the amount of **rising_a** contours increases to 41 % and **rising_w** contours emerge as well (17 %). With the amount of **falling-rising** contours staying about the same (7 %), this leaves 35 % **falling** contours in this second round of task 3.

The increase of rising contours (this includes **falling-rising**, **rising_a**, and **rising_w** contours) from round 1 to round 2 of task 3 was evaluated by fitting a generalized linear mixed-effects model (GLMM) with the binary response **falling/rising** as dependent variable and including the contrast coded factors **task** (round1-round2) and **gender** (female-male). The most complex model allowing a non-singular fit [15] includes random intercepts for **user** and random slopes for **task by user**.[1] The factor **task** is a significant predictor of the dependent variable with the following parameters: estimate (log-odds) $= -4.91$, SE $= 1.31$, $z = -3.75$, $p < 0.001$. The factor **gender** does not explain any variance in the data.

3.2. Segments

Of the 20 users participating in the experiment, 10 had a preference for the [ɪç] variant (8 female) and 10 for the [ɪk] variant (8 female) in the baseline task; 12 users had a preference for [ɛː] (10 female) and 8 for [eː] (6 female).

Figure 2 shows the results of the [ɪç] vs. [ɪk] contrast. In 90 % of all baseline task instances, the users produce a **different variant** of the target contrast than they hear from Mirabella in the map task. The remaining 10 % are cases where users utter the dispreferred variant in the baseline task, hence the **same variant** as Mirabella. Both categories are evenly distributed over the users who prefer [ɪç] and those who prefer [ɪk] (different: 45 % each group; same: 5 % each group). In the map task, the amount of dispreferred variants uttered by the users increases by 27 % to a total of 37 %. The distribution over the two preference groups is again very even with 19 % of the instances stemming from the [ɪç] group and 18 % from the [ɪk] group.

[1] Statistical tests are carried out in RStudio (v1.1.463) [16] with R (v3.5.2) [17] using the packages lme4 (v1.1.-21) [18] and lmerTest (v3.1-0) [19].

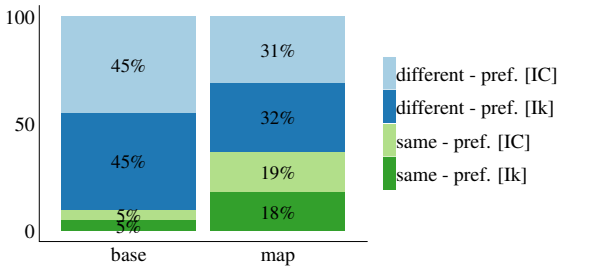

Figure 2: *Instances where the user and Mirabella produce **different variants** or the **same variant** of the target contrast [ɪç] vs. [ɪk] during task 1 (base) and task 4 (map). These two main categories are further divided by the users' overall preference for either [ɪç] or [ɪk].*

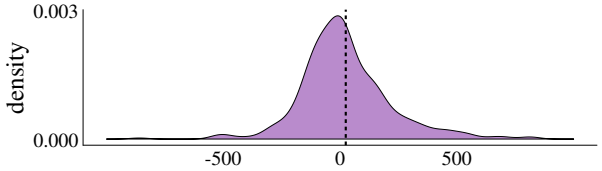

Figure 3: *Difference in Euclidean distance (in Hz) in the F1-F2 space between user realizations of ⟨-ä-⟩ and the respective realizations by Mirabella in baseline compared to map task. The mean of 24.5 is indicated by the dashed line. Positive values indicate convergence, negative values divergence.*

Figure 3 shows the distribution of the difference in Euclidean distance (*dDist*) in the F1-F2 space between user realizations of long, stressed ⟨-ä-⟩ and the respective realizations by Mirabella in the baseline task compared to the map task. The distribution has a mean of 24.5 which is positive and therefore tends towards convergence.

However, fitting a linear mixed-effects model with **dDist** as dependent variable, including the contrast coded factors **gender** (female-male) and **preference** ([ɛː]-[eː]), as well as random intercepts for **user** and **target word** and random slopes for **preference by target word**, reveals that the mean does not differ significantly from zero (estimate $= 19.24$, SE $= 31.82$, df $= 17.69, t = 0.61$, $p = 0.55$). The factors **gender** and **preference** do not explain any variance in the data.

4. Discussion

We found that the 20 users who took part in the WOz experiment, as a group, showed substantial convergence to the speech output of Mirabella with respect to the intonation of constituent questions and the realization of the allophonic contrast [ɪç] vs. [ɪk]. The effect was more pronounced for the intonation feature

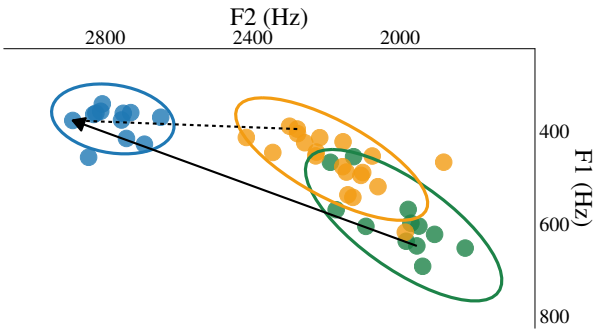

Figure 4: *Realizations of ⟨-ä-⟩ by user f04 during task 1 (**base**) and task 4 (**map**) relative to Mirabella's productions of [ɛː] during task 4 (**map**). The arrows illustrate the Euclidean distance for the target word **ähnlich** as an example. The ellipses indicate the 95 % confidence interval.*

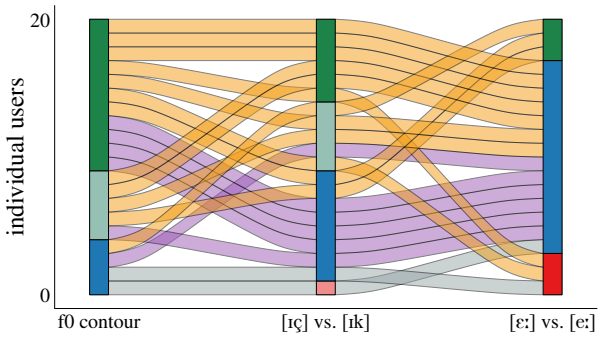

Figure 5: *Accommodation behavior for the 20 individual users over the three examined features. The colors code **substantial conv.**, **moderate conv.**, **maintenance**, **moderate div.**, and **substantial div.**. Some users converge with respect to **two features**, some only for **one feature**, and some do **not converge** at all.*

(increase of rising contours by 52 %) than for the allophonic contrast (increase of dispreferred variants by 27 %).

The convergence effect for the [ɪç] vs. [ɪk] contrast is lower than previously found in a shadowing experiment with natural and synthetic stimuli, where participants adopted the variant they heard from a model voice in about 37 % of the cases [20]. However, it should be noted that speech shadowing entails direct repetition of an utterance, while the users formulated entirely new utterances in the map task at hand. Therefore the slightly lower convergence effect is not surprising and might still be even more meaningful, since the repetition component is much reduced.

The convergence effect for the question intonation was mainly driven by an increase (34 %) of rising contours with a nuclear pitch accent on the respective animal. Only 17 % of all questions in round 2 of task 3 were produced with rising contours and a nuclear pitch accent on the interrogative pronoun *wo*, which is the way Mirabella uttered her question in this subtask. This suggests that users are primarily receptive for the overall rising contour and to a lesser degree also shift the nuclear pitch accent.

For the allophonic contrast [ɛː] vs. [eː], the result for the entire user group was overall maintaining behavior. This is unexpected, since a convergence effect was found for the natural stimuli in the above-mentioned shadowing experiment and Mirabella's utterances were natural speech too.

Taking a closer look at the behavior of the individual users reveals that convergence and maintenance occur for all three features, whereas divergence occurs only for the allophonic contrasts. Even for the [ɛː] vs. [eː] contrast, where no effect was found at the group level, three participants moved their realizations of ⟨-ä-⟩ significantly towards those of Mirabella (evaluated by two-sided one-sample Wilcoxon signed-rank test with $\alpha = 0.05$ for each participant's distribution of *dDist*). Figure 4 illustrates one of these cases. Another three participants significantly diverged from Mirabella. It has to be noted here that we are evaluating proximity in the F1-F2 space and not a categorical change between [ɛː] and [eː], which would certainly be a valuable extension of the present analysis.

The question arises whether users always converge, maintain, or diverge for all three examined features jointly. Figure 5 presents an overview of the individual user behavior. The degree of accommodation was categorized as follows:

Substantial convergence at least seven instances of convergence for [ɪç] vs. [ɪk], at least five instances of *rising_a* or *rising_w*, or a significantly positive *dDist* for [ɛː] vs. [eː].

Moderate convergence at least two instances of convergence for [ɪç] vs. [ɪk] and question intonation. This category is not available for [ɛː] vs. [eː].

Maintenance up to one case of convergence or divergence for [ɪç] vs. [ɪk] and question intonation, as well as a *dDist* not significantly different from 0 for [ɛː] vs. [eː].

Moderate and substantial divergence see convergence.

There are three top convergers for the [ɛː] vs. [eː] contrast, six top convergers for the [ɪç] vs. [ɪk] contrast, and eleven top convergers for the question intonation. The users maximally converge to Mirabella for two features jointly (12 users, see orange cases in Figure 5). Only two users do not show any convergence, yet diverge with respect to one feature. Furthermore, maintaining behavior for one feature is found in nine users, and for two features in eight users. Divergence, eventually, occurs in a total of four individual cases.

5. Conclusion

We conducted a WOz experiment with 20 native German speakers to investigate phonetic accommodation on the part of the user in HCI. The participants interacted with the simulated SDS Mirabella to jointly solve a number of tasks embedded in a CALL scenario. The tasks were designed to give room for accommodative behavior with respect to the intonation of wh-questions and two allophonic contrasts. On the group level, we found convergence to Mirabella for the question intonation and the allophonic contrast [ɪç] vs. [ɪk] with a stronger effect for the former, yet overall maintenance of the preferred variant for the [ɛː] vs. [eː] contrast. On the level of individual users, we found cases of convergence and maintenance for all three examined features, as well as occasional divergence. Overall, users accommodated to the speech output of the simulated SDS, but to highly individual degrees, and the performance in one feature did not seem to predict performance in another feature.

We are planning to extend the user group to non-native speakers of German, and apply synthetic speech instead of natural recordings in Mirabellas's utterances.

6. Acknowledgments

This research was funded in part by the German Research Foundation (DFG) under grants MO 597/6-2 and STE 2363/1-2. We thank Jens Neuerburg (annotations) and Katie Ann Dunfield (recordings).

7. References

[1] J. Pardo, "On phonetic convergence during conversational interaction", *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006. DOI: 10.1121/1.2178720.

[2] G. Bailly and A. Lelong, "Speech dominoes and phonetic convergence", in *Interspeech*, Makuhari, 2010, pp. 1153–1156. [Online]. Available: https://www.isca-speech.org/archive/interspeech_2010/i10_1153.html.

[3] N. Lewandowski, "Talent in nonnative phonetic convergence", PhD thesis, Universität Stuttgart, 2012. DOI: 10.18419/opus-2858.

[4] M. Babel, G. McGuire, S. Walters, and A. Nicholls, "Novelty and social preference in phonetic accommodation", *Laboratory Phonology*, vol. 5, no. 1, pp. 123–150, 2014. DOI: 10.1515/lp-2014-0006.

[5] H. Giles, N. Coupland, and J. Coupland, "Accommodation theory: Communication, context, and consequence", in *Contexts of Accommodation: Developments in Applied Sociolinguistics*, H. Giles, J. Coupland, and N. Coupland, Eds., Cambridge University Press, 1991, pp. 1–68. DOI: 10.1017/CBO9780511663673.001.

[6] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors", in *ACM SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, 1994, pp. 72–78. DOI: 10.1145/191666.191703.

[7] H. P. Branigan, M. J. Pickering, J. Pearson, and J. F. McLean, "Linguistic alignment between people and computers", *Journal of Pragmatics*, vol. 42, no. 9, pp. 2355–2368, 2010. DOI: 10.1016/j.pragma.2009.12.012.

[8] L. Bell, J. Gustafson, and M. Heldner, "Prosodic adaptation in human-computer interaction", in *International Congress of Phonetic Sciences (ICPhS)*, Barcelona, 2003, pp. 2453–2456. [Online]. Available: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_2453.html.

[9] S. Oviatt, C. Darves, and R. Coulston, "Toward adaptive conversational interfaces: Modeling speech convergence with animated personas", *ACM Transactions on Computer-Human Interaction*, vol. 11, no. 3, pp. 300–328, 2004. DOI: 10.1145/1017494.1017498.

[10] L. Gauder, M. Reartes, R. H. Gálvez, Š. Beňuš, and A. Gravano, "Testing the effects of acoustic/prosodic entrainment on user behavior at the dialog-act level", in *Speech Prosody*, Poznań, 2018, pp. 374–378. DOI: 10.21437/SpeechProsody.2018-76.

[11] I. Gessinger, B. Möbius, N. Fakhar, E. Raveh, and I. Steiner, "A Wizard-of-Oz experiment to study phonetic accommodation in human-computer interaction", in *International Congress of Phonetic Sciences (ICPhS)*, in press, Melbourne, 2019.

[12] D. Wochner, J. Schlegel, N. Dehé, and B. Braun, "The prosodic marking of rhetorical questions in German", in *Interspeech*, Dresden, 2015, pp. 987–991. [Online]. Available: https://www.isca-speech.org/archive/interspeech_2015/i15_0987.html.

[13] M. Grice and S. Baumann, "Deutsche intonation und GToBI", *Linguistische Berichte*, pp. 267–298, 2002.

[14] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*, 2017. [Online]. Available: https://www.praat.org/.

[15] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal", *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013. DOI: 10.1016/j.jml.2012.11.001.

[16] RStudio Team, *Rstudio: Integrated development environment for r*, RStudio, Inc., Boston, MA, 2016. [Online]. Available: https://www.rstudio.com/.

[17] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.r-project.org/.

[18] 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. DOI: 10.18637/jss.v067.i01.

[19] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models", *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017. DOI: 10.18637/jss.v082.i13.

[20] I. Gessinger, E. Raveh, S. Le Maguer, B. Möbius, and I. Steiner, "Shadowing synthesized speech – segmental analysis of phonetic convergence", in *Interspeech*, Stockholm, 2017, pp. 3797–3801. DOI: 10.21437/Interspeech.2017-1433.