INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Phonetic Accommodation of L2 German Speakers to the Virtual Language Learning Tutor Mirabella

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

The present paper compares phonetic accommodation of L1 French speakers in interaction with the simulated virtual language learning tutor for German, Mirabella, to that of L1 German speakers from a previous study. In a question-and-answer exchange, the L1 French speakers adapted the intonation contours of wh-questions as falling or rising according to the variant produced by Mirabella. However, they were not sensitive to a change of the nuclear pitch accent placement. In a map task, the L1 French speakers increased the number of dispreferred variants for the allophonic contrast [ɪç] vs. [ɪk] in the word ending ⟨-ig⟩ when Mirabella used this variant. For the contrast [ɛː] vs. [eː] as a realization of stressed ⟨-ä-⟩, such a convergence effect was not found. Overall, the non-native speakers showed a similar degree of accommodative behavior towards Mirabella as the L1 German speakers. This suggests that incidental inductive learning through accommodation is possible. However, phenomena of the target language that deviate too radically from the native pattern seem to require more explicit training.

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

1. Introduction

Computer-assisted language learning (CALL) offers a low-threshold opportunity to start acquiring a foreign language. There is a wide range of options on the market, from simple, free applications to sophisticated learning programs. On the one hand, it is often argued that a CALL application cannot fully replace a human teacher, especially when it comes to phonetic aspects of communication [1]. On the other hand, progress in all areas of the human-computer interaction (HCI) domain, particularly automatic speech recognition and text-to-speech synthesis, enables increasingly successful spoken communication between human learners and virtual teachers. Furthermore, we believe that the demand to learn a foreign language can exceed the supply of human teachers, and virtual alternatives have to be applied, for example in the context of mass migration. It is therefore both promising and necessary to further develop CALL applications and improve the learning experience.

Phonetic accommodation describes the phenomenon that human interlocutors adapt their speech to each other during conversational interaction by either becoming more similar (converging) or dissimilar (diverging) [2, 3, 4]. Apart from an explanatory approach that suggests automatic perception-production integration as the reason for convergence [5, 6], it is assumed that accommodation serves to regulate social distance, with convergence increasing proximity and divergence decreasing it [7, 8]. In line with the social approach, the direction and extent of accommodation have been found to depend on factors such as the attitude or the hierarchical relationship towards an interlocutor. For example, it has been shown that an increase in the likability of a conversational partner led to a stronger convergence effect for vowel quality [9] and fundamental frequency [10]. However, Schweitzer *et al.* [11] report that a decrease in likability promoted both convergence and divergence with respect to pitch accent realization. Results by Gregory and Webster [12], analyzing long-term average spectra, suggest that speakers on the lower end of the hierarchy, or in a less dominant role, converge to the hierarchically higher or more dominant interlocutor.

A language learning scenario may benefit from this dynamic process of phonetic adaptation. The teacher detects incorrect pronunciation on the part of the learner and intentionally diverges from it. For a continuous feature, e.g., vowel quality, this may imply producing a more extreme version, whereas for a categorical feature, e.g., allophonic variation, this may mean emphasizing the preferred version. The learner then converges to the teacher's pronunciation, especially if the teacher is perceived as likable and hierarchically superior. While this is a conceivable scenario for human-human interaction, the question arises as to whether such dynamic phonetic adaptation can also occur in HCI.

On the computer side, the ability to phonetically accommodate is not yet available; suggestions for possible implementations are being developed [13, 14]. On the human side, a relevant factor in this context is the perception of the virtual tutor as an actual social counterpart and not merely as a tool. It was observed that computers can indeed be perceived as social actors and that people show social behavior towards them [15, 16, 17], but this does not necessarily apply to every person equally [18]. The design of a virtual tutor can influence the degree of personification, for example by having a name, being represented by an avatar or speaking with a natural-sounding voice.

A small number of studies have explored whether humans accommodate to the output of spoken dialogue systems (SDSs) using embodied virtual agents [19, 20, 21, 22]. They applied the Wizard-of-Oz (WOz) method to simulate SDSs, which seem autonomous to the user, but are in fact controlled by the experimenter. They found that the users of these systems showed converging behavior on global acoustic-prosodic features, such as F_0, intensity, and speaking rate.

The present paper examines phonetic accommodation in a CALL scenario with respect to more locally anchored features, which have variations that are accepted in Standard German: the placement of the nuclear pitch accent in wh-questions, the final intonation contour following this nuclear accent, and the German allophone pairs [ɪç] vs. [ɪk] occurring in the word ending ⟨-ig⟩ and [ɛː] vs. [eː] as a realization of the long vowel ⟨-ä-⟩. We applied the WOz method to create a dynamic conversational exchange between the participants and a virtual tutor for learning German called *Mirabella* [23], with direct control over her

http://dx.doi.org/10.21437/Interspeech.2020-2701

speech output. We tested L2 German speakers and compared their results with those of L1 German speakers who showed accommodative behavior to Mirabella in a previous study [24]. On the one hand, it is conceivable that non-native speakers show more adaptation than native speakers because they are less confident in their own pronunciation and Mirabella, as a perceived native speaker of the target language, is hierarchically above them. On the other hand, it is possible that they have greater difficulty in perceiving the phonetic detail and implementing it in their production, and therefore accommodate less.

2. Method

Native speakers of French ($n = 11$; 5 female, 6 male; mean age 25 years, age range 16 to 53 years) took part in the experiment which was presented to them as a test run for the virtual language learning tutor Mirabella. The test was about "learning German"; the subject of pronunciation was not mentioned at any point. The participants were students or employees of educational institutions in Saarbrücken, Germany, and came from different regions of France ($n = 10$) and Cameroon ($n = 1$). All participants indicated French as their sole or dominant native language. The participants spoke 2 to 4 foreign languages; their command of German ranged from *B2: upper intermediate* ($n = 2$) to *C1: advanced* ($n = 9$) according to the Common European Framework of Reference for Languages (CEFRL).

During the experiment, the participants only interacted with Mirabella's voice; she was not represented by an embodied virtual agent. The interaction was supported by visualization of the tasks on a screen [23]. Mirabella explained the tasks to the participants and took part in them, either by taking turns in a question-and-answer (Q&A) exchange with the participants or by providing missing information to the participants in a map task. Mirabella's utterances were pre-recorded by a female native speaker of German (aged 26 years) and manually played back to the participants by the experimenter.

The interactions with Mirabella comprised four tasks, lasted about 40 minutes, and were recorded at a sampling rate of 48 kHz using a stationary cardioid microphone in a sound-attenuated booth with the participants sitting in front of a screen.

2.1. Question intonation

Baseline: The participants formulated five constituent wh-questions whose components were given as fragments. The questions were answered by Mirabella.

Q&A: Mirabella and the participants took turns asking and answering each other about ten animals hiding in ten houses on the screen. In the present analysis, we only examined the questions, which were formulated: *Wo hat sich ⟨the animal⟩ versteckt?* ("Where did ⟨the animal⟩ hide?") The task was divided into two rounds of 20 turns each. The order in which Mirabella and the participants asked for the animals on the screen was free. In the first round, Mirabella produced all questions with a nuclear pitch accent on the animal, followed by a final F_0 fall, which is the expected intonation contour for German wh-questions [25]. In the second round, she produced a nuclear pitch accent on the question word *wo*, followed by a final high F_0 rise, which in German occurs mainly in echo questions [26], i.e., when an answer was not understood and the question is asked again. In the context of the present Q&A it is unexpected, but not pragmatically wrong, to ask with rising intonation for the location of the animals in the second round [24].

Prediction: French wh-questions are generally realized with a final F_0 fall, but rising contours are possible as well; as in German, such rising contours mainly occur in echo questions [27, 28]. We expected mainly falling intonation contours for the questions in the baseline task and the first round of the Q&A, and an increase of rising contours from the first to the second round of the Q&A.

Analysis: Two trained phoneticians marked the position of the nuclear pitch accent in 284 wh-questions and classified the final intonation contours following this nuclear accent as *falling*, *rising*, or other occurring types.

2.2. Allophonic variation

Baseline: The participants named pictures and translated French adjectives to German by uttering them in the carrier sentence: *Das Wort ⟨item⟩ kenne ich.* ("The word ⟨item⟩ is known to me.") A subset of these items contained the allophonic contrasts [ɪç] vs. [ɪk] (e.g., in *traurig*, Example (a)) or [ɛː] vs. [eː] (e.g., in *Jäger*, Example (b)). While [ɪç] and [ɛː] are the codified Standard German variants, [ɪk] and [eː] are regional variants, which are however also perceived as (almost) non-dialectal [29]. The individual realizations were perceptually categorized by the experimenter to determine the participants' preferred variants.

Map task: The participants described all objects on a map from leaving a house until reaching a destination in two-part statements like:

(a) *Ich gehe an dem Fisch vorbei. Der Fisch ist* ***traurig****.*
"I am walking past the fish. The fish is **sad**."

(b) *Ich gehe um den* ***Jäger*** *herum. Der* ***Jäger*** *ist gesund.*
"I am walking around the **hunter**. The **hunter** is healthy."

Some of the items were hidden behind boxes. Mirabella provided the missing information while using the participants' dispreferred variant of the respective allophonic contrast – thus effectively diverging from them. Given this information, the participants formulated the required two-part statement. The task consisted of four maps and contained a total of 12 occurrences per allophonic contrast.

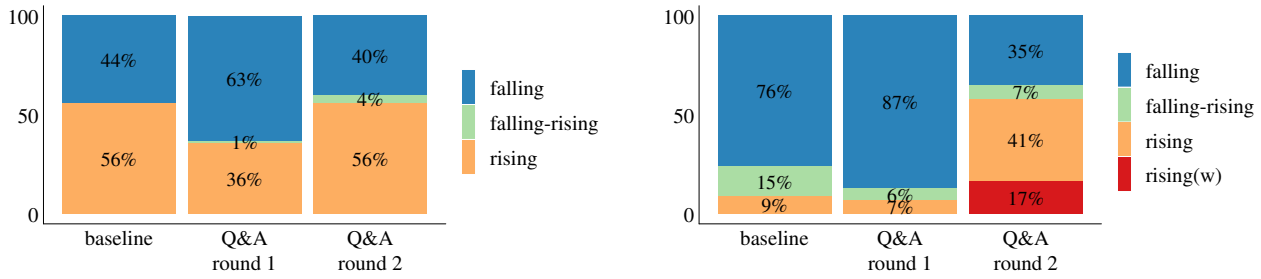
Prediction: For the [ɪç] vs. [ɪk] contrast we expected an increase of the dispreferred variant during the map task compared to the baseline task. With regard to the [ɛː] vs. [eː] contrast we only found accommodation effects for a few individual native speakers of German [cf. 24]. Therefore, we did not expect a group effect for the non-native speakers, but possibly effects at the level of individual speakers.

Analysis: The 286 realizations of the word ending ⟨-ig⟩ were perceptually classified by a trained phonetician as belonging to the fricative or plosive category and as being the same as or a different variant than the one produced by Mirabella. The fricative category included variants of [ɪç] such as [ɪʃ].

For the 348 realizations of long, stressed ⟨-ä-⟩, the first and second formants were measured at their midpoint using Praat's [30] Burg algorithm. First, the Euclidean distance in the F1-F2 space between each participant realization and the corresponding realization by Mirabella was calculated. Then, the difference in Euclidean distance between baseline task and map task was determined. The latter is zero if the participants did not shift their productions in the F1-F2 space (maintenance), above zero if they shifted their productions in the direction of Mirabella (convergence), and below zero if they shifted their productions away from Mirabella (divergence).

3. Results

In a questionnaire after the experiment, the participants rated Mirabella on 5-point scales as very likable (*unpleasant* to *very*



(a) *Non-native speakers of German ($n = 11$).* (b) *Native speakers of German ($n = 20$); modified from [24].*

Figure 1: *Percentages of questions realized with **falling**, **falling-rising**, or **rising** intonation contour during the baseline task and the two rounds of the Q&A game. Cases of **rising(w)** indicate a shift of the nuclear pitch accent to the interrogative word.*

likable – mean: 4.6), very competent (*incompetent* to *very competent* – mean: 4.7) and very well intelligible (*badly* to *very well* – mean: 5). They also considered Mirabella's reaction time to be appropriate (*too slow* to *too fast* – mean: 3.3).

3.1. Question intonation

Figure 1 shows the results of the question intonation analysis for the non-native speakers (1a) and compares them with the results of the native speakers from [24] (1b). We found *falling*, *falling-rising*, and *rising* contours [26] in the data of the non-native speakers. While *falling* contours were predominant in the baseline productions of the native speakers, the non-native speakers produced 56 % *rising* and only 44 % *falling* contours in the same task. Like the German group, the French speakers produced more *falling* contours (63 %) when interacting with Mirabella in the first round of the Q&A. However, they still produced a substantial amount of *rising* (36 %) and some *falling-rising* (1 %) contours, as well. In the second round of the Q&A, where Mirabella produced *rising(w)* contours with a nuclear pitch accent on the interrogative word, the amount of *rising* (56 %) and *falling-rising* (4 %) contours in the French group increased again. Unlike the German speakers, however, the French speakers did not produce any *rising(w)* contours.

The increase of rising contours (i.e., *falling-rising* and *rising*) from the first to the second round of the Q&A was evaluated by fitting a generalized linear mixed-effects model (GLMM) to the data of the non-native speakers.[1] Although including the contrast coded factor TASK (round1-round2) improved the fit of the model, it was not a significant predictor of the contour type (estimate (log-odds) $= -2.16$, SE $= 2.57$, $z = -0.84$, $p = 0.4$). The model included random intercepts for PARTICIPANT and ITEM, as well as by-participant random slopes for TASK. The increase of rising contours was significant for the native speaker group [24].

Two further points are noteworthy in the non-native speaker data: First, the questions were not always produced as one single intonational phrase, but some participants had a tendency to produce the final part of the question, *versteckt* ("hidden"), separately. This occurred in 12 % of all questions in the first Q&A round, but only 6 % in the second Q&A round. Second, the nuclear pitch accent was not always realized on the respective animal, but sometimes on the ultima of *versteckt* (which coincides with the lexical stress in German). While this can partly be an effect of the unusual phrasing mentioned above, it occurred more frequently, namely in 33 % of all questions in the first Q&A round and 20 % in the second Q&A round. Whereas the decrease in cases of unusual phrasing was not significant in a GLMM with random intercepts for PARTICIPANT and ITEM (estimate (log-odds) $= 0.46$, SE $= 0.28$, $z = 1.66$, $p = 0.1$), the increase of nuclear pitch accents on the respective animal in the second round of the Q&A was significant in an equivalent model (estimate (log-odds) $= 0.51$, SE $= 0.19$, $z = 2.68$, $p < 0.01$).

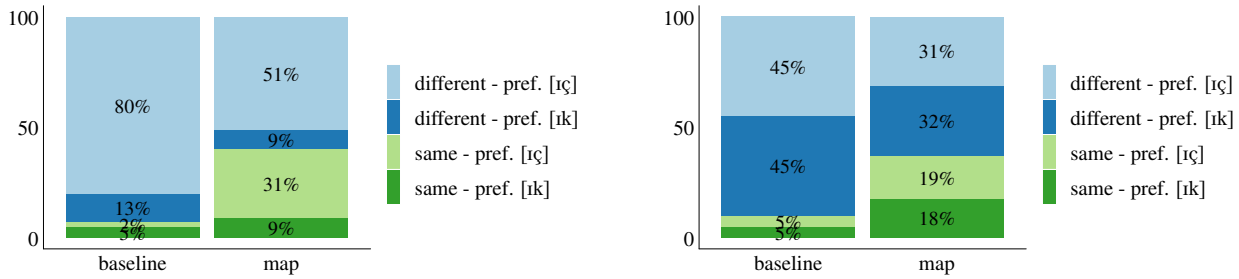
3.2. Allophonic variation

Figure 2 shows the results of the [ɪç] vs. [ɪk] contrast for the non-native speakers (2a) and compares them with the results of the native speakers from [24] (2b). In 93 % of all baseline task instances, the French participants produced a *different variant* of the target contrast than they heard from Mirabella in the map task. The remaining 7 % are cases where the participants uttered the dispreferred variant in the baseline task, hence the *same variant* as Mirabella. In the map task, the amount of dispreferred variants uttered by the non-native speakers increased by 33 % to a total of 40 %. The baseline distribution and the accommodative effect in the map task is nearly identical to the native speakers, with the only difference that the majority of the French participants ($n = 9$) had a baseline preference for [ɪç], while the German participants were equally distributed between the two preference groups.

The increase of dispreferred variants was evaluated by fitting a GLMM. The most complex model allowing a non-singular fit [33] included the contrast coded factors TASK (base-map) and PREFERENCE ([ɪç]-[ɪk]), random intercepts for PARTICIPANTS and ITEM, as well as by-participant random slopes for TASK. Both the factor TASK (estimate (log-odds) $= -1.02$, SE $= 0.36$, $z = -2.80$, $p < 0.01$) and the factor PREFERENCE (estimate (log-odds) $= -1.45$, SE $= 0.49$, $z = -2.94$, $p < 0.01$) were significant predictors for *different/same*.

The distribution of the difference in Euclidean distance (dDist) in the F1-F2 space between the non-native speakers' realizations of long, stressed ⟨-ä-⟩ and the respective realizations by Mirabella in the baseline task compared to the map task had a mean of -12 which is negative and therefore tends towards divergence. 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 PARTICIPANT and ITEM, revealed that the mean does not differ significantly from zero (estimate $= -23.93$, SE $= 24.83$, df $= 12.99, t = -0.96$, $p = 0.35$). The factors GENDER and PREFERENCE did not explain any variance in the data. This was also the case for the native speakers of German [24]. Evaluating each participant's individual distribu-

[1] Statistical tests were carried out using the R packages lme4 (v1.1.-21) [31] and lmerTest (v3.1-0) [32].



(a) *Non-native speakers of German ($n = 11$).*

(b) *Native speakers of German ($n = 20$); from [24].*

Figure 2: *Instances of the participants and Mirabella producing **different variants** or the **same variant** of the target contrast [ɪç] vs. [ɪk] during baseline and map tasks. The two categories are further divided by the participants' overall preference for either [ɪç] or [ɪk].*

tion of *dDist* by a two-sided one-sample Wilcoxon signed-rank test ($\alpha = 0.05$) yielded a significantly positive deviation from zero for participant *m07*, hinting at convergence, and a significantly negative deviation from zero for participant *m05*, hinting at divergence. An additional kernel density based global two-sample comparison test for two-dimensional data ($\alpha = 0.05$), however, showed that for neither of the two participants the baseline vowels differed significantly from their own map task vowels, which means that they did not substantially change their own vowel distribution. Following the same approach for the native speakers of German, which was not a part of the previous study [24], revealed two cases of convergence towards, and two cases of divergence from, Mirabella.

4. Discussion and conclusion

We compared the accommodative behavior of 11 native speakers of French in interaction with a virtual tutor for learning German called Mirabella with that of 20 native speakers of German from a previous study [24]. When questioned after the experiment, the L1 French speakers had no suspicion that Mirabella was testing particular pronunciation-related phenomena. One assumption was that the test was about how well artificial intelligence can understand non-native speakers, which indicates that Mirabella was perceived as an intelligent system.

Regarding the question intonation, the L1 French speakers had no particular preference for *falling* or *rising* contours in their baseline productions. As it was the case for the L1 German speakers, the number of *falling* contours increased as Mirabella produced *falling* contours in the first Q&A round, and the number of *rising* contours increased as she produced *rising* contours in the second Q&A round. However, the increase from first to second round was not significant for the L1 French speakers.

In the case of the German group, Mirabella's shift of the nuclear pitch accent to the question word *wo* ("where") was adopted by some speakers. This never happened in the French group. It can be assumed that the metrical pattern of French, which uses relatively small accentual phrases and has an obligatory phrase-final accent [27], contradicts the realization of an initial nuclear pitch accent considerably and therefore even advanced learners (here: CEFRL B2/C1) do not adopt this pattern. In this case a limit of accommodation may have been reached. To emphasize the question word and still follow the native pattern, French would favor a syntactic variation in combination with a *rising* contour, namely: *L'animal se cache où?* ("The animal is hiding where?") [28].

The decrease in the production of unusual phrasing in the second Q&A round and the significant increase in the number of nuclear pitch accents placed on the animal instead of the ultima of *versteckt* ("hidden"), may be interpreted either as a reduction of insecurity in the interaction with Mirabella on the part of the L1 French speakers, or as accommodation of their own native pattern to that of the foreign language.

For the allophonic contrast [ɪç] vs. [ɪk], both L1 French and L1 German speakers exhibited a significant increase in dispreferred variants when communicating with Mirabella in a map task, which means that the participants converged to her. The majority of the L1 French speakers had a baseline preference for the fricative variant, while both variants were equally common among the L1 German speakers. This may be due to the fact that the non-native speakers learned German through formal instruction and [ɪç] is the codified Standard German variant.

Finally, for the allophonic contrast [ɛː] vs. [eː] neither the L1 German speakers nor the L1 French speakers adapted to Mirabella at the group level. While there were some individual cases of convergence and divergence in the German group, the non-native speakers do not seem to have adapted their vowel quality to Mirabella. The present analysis of the vowel quality is stricter than that of the [ɪç] vs. [ɪk] contrast, since the distribution is considered as a whole and individual cases of accommodation are thus not taken into account. A perceptual analysis would be a valuable extension to the present results.

For all three examined features, the non-native speakers of German behave similarly to the native speakers, but exhibit specific patterns especially in the case of the question intonation. Where the L1 French speakers were adapting their speech to Mirabella, it was in the form of convergence. This is consistent with the assumption that convergence occurs in the direction of the hierarchically superior interlocutor, since Mirabella has a model function as a native speaker of the target language and also provides information for the participants to solve the tasks. Furthermore, the participants rated Mirabella as very likable and competent, showing that they had a positive attitude towards her, which can also be conducive to convergence.

We conclude that non-native speakers interacting with a virtual language learning tutor show a similar degree of accommodative behavior towards the latter as do native speakers, which enables incidental inductive learning [34]. However, structural phonological elements of the target language that deviate too radically from the native pattern seem to require more explicit training.

5. Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID MO 597/6-2 and STE 2363/1-2. We thank Jens Neuerburg for assistance with recordings and annotations.

6. References

[1] J. Lee, J. Jang, and L. Plonsky, "The effectiveness of second language pronunciation instruction: A meta-analysis", *Applied Linguistics*, vol. 36, no. 3, pp. 345–366, 2014.

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

[3] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions", in *Interspeech*, Florence, 2011, pp. 3081–3084.

[4] N. Lewandowski, "Talent in nonnative phonetic convergence", PhD thesis, Universität Stuttgart, 2012.

[5] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue", *Behavioral and Brain Sciences*, vol. 27, no. 2, pp. 169–190, 2004.

[6] ——, "An integrated theory of language production and comprehension", *Behavioral and Brain Sciences*, vol. 36, no. 4, pp. 329–347, 2013.

[7] 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.

[8] 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., Wiley, 2001, pp. 33–56.

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

[10] J. Michalsky and H. Schoormann, "Pitch convergence as an effect of perceived attractiveness and likability", in *Interspeech*, Stockholm, 2017, pp. 2253–2256.

[11] K. Schweitzer, M. Walsh, and A. Schweitzer, "To see or not to see: Interlocutor visibility and likeability influence convergence in intonation", in *Interspeech*, Stockholm, 2017, pp. 919–923.

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

[13] R. Levitan, Š. Beňuš, R. H. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, "Implementing acoustic-prosodic entrainment in a conversational avatar", in *Interspeech*, San Francisco, CA, 2016, pp. 1166–1170.

[14] E. Raveh, I. Steiner, and B. Möbius, "A computational model for phonetically responsive spoken dialogue systems", in *Interspeech*, Stockholm, 2017, pp. 884–888.

[15] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors", in *SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1994, pp. 72–78.

[16] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people and places", *Computers & Mathematics with Applications*, vol. 33, no. 5, p. 128, 1996.

[17] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers", *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[18] M. K. Lee, S. Kiesler, and J. Forlizzi, "Receptionist or information kiosk: How do people talk with a robot?", in *ACM Conference on Computer Supported Cooperative Work*, 2010, pp. 31–40.

[19] 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.

[20] 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, pp. 300–328, 2004.

[21] L. Staum Casasanto, K. Jasmin, and D. Casasanto, "Virtually accommodating: Speech rate accommodation to a virtual interlocutor", in *Conference of the Cognitive Science Society*, Cognitive Science Society, 2010, pp. 127–132.

[22] T. Gijssels, L. Staum Casasanto, K. Jasmin, P. Hagoort, and D. Casasanto, "Speech accommodation without priming: The case of pitch", *Discourse Processes*, vol. 53, no. 4, pp. 233–251, 2016.

[23] 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)*, Melbourne, 2019, pp. 1475–1479.

[24] I. Gessinger, B. Möbius, B. Andreeva, E. Raveh, and I. Steiner, "Phonetic accommodation in a Wizard-of-Oz experiment: Intonation and segments", in *Interspeech*, Graz, 2019, pp. 301–305.

[25] D. Wochner, J. Schlegel, N. Dehé, and B. Braun, "The prosodic marking of rhetorical questions in German", in *Interspeech*, Dresden, 2015, pp. 987–991.

[26] M. Grice and S. Baumann, "Deutsche Intonation und GToBl", *Linguistische Berichte*, pp. 267–298, 2002.

[27] A. Di Cristo, "Intonation in French", in *Intonation systems: A survey of twenty languages*, D. Hirst and A. Di Cristo, Eds., Cambridge University Press, 1998, ch. 11, pp. 195–218.

[28] E. Delais-Roussarie, B. Post, M. Avanzi, C. Buthke, A. Di Cristo, I. Feldhausen, S.-A. Jun, P. Martin, T. Meisenburg, A. Rialland, R. Sichel-Bazin, and H.-Y. Yoo, "Intonational phonology of French: Developing a ToBI system for French", in *Intonation in Romance*, S. Frota and P. Prieto, Eds., Oxford University Press, 2015, pp. 63–100.

[29] C. Kiesewalter, *Zur subjektiven Dialektalität regiolektaler Aussprachemerkmale des Deutschen*, ser. Zeitschrift für Dialektologie und Linguistik, Beihefte 179. Franz Steiner Verlag, 2019.

[30] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer [computer program]*, Version 6.1.06, retrieved 8 November 2019 from http://www.praat.org/, 2019.

[31] 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.

[32] 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.

[33] 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.

[34] J. N. Williams, "Implicit learning in second language acquisition", in *The new handbook of second language acquisition*, W. C. Ritchie and T. K. Bhatia, Eds., Emerald, 2009, ch. 14, pp. 319–353.