Why so cold and distant? Effects of inter-turn gap durations on observers' attributions of interpersonal stance

Stefan Blohm and Mathias Barthel
Institute for the German language (IDS)
Mannheim, Germany
blohm@ids-mannheim.de -- barthel@ids-mannheim.de

Abstract

Gaps between turns of talk in fluent conversation are typically very short, while longer gaps often indicate 'trouble' (e.g., upcoming disagreement with a prior assessment). Listeners use this timing cue to infer respondents' stance, e.g., their willingness to comply with a request. We tested whether gap durations also influence attributions of interpersonal affective stances ("cold and distant reaction") and whether listeners flexibly adjust their temporal threshold of tolerance to speaker characteristics (native language: L1 vs. L2).

We constructed short German dialogues that contained critical question-response pairs; we manipulated the duration of inter-turn gaps and whether the respondent was a native vs. non-native speaker of German. Participants listened to these critical question-response pairs and judged "how cold and distant the respondent reacts". Bayesian mixed-effects regression revealed that longer-than-average gaps increased attributions of interpersonal distance as well as participants' response times, whereas respondents' nativeness affected participants' judgments but not their response times. Our results did not confirm the predicted interaction of gap duration and nativeness, thus failing to support the idea that these attributional inferences are delayed or attenuated for non-native respondents. Among the phonetic parameters we controlled for, we find that the duration of the response token affected judgments of interpersonal distance, presumably reflecting that very short responses convey curtness.

We conclude that listeners use the duration of inter-turn silence and other phonetic parameters as well as general speaker characteristics as heuristic cues to respondents' interpersonal affective stance.

1 Introduction

In fluent conversation, interlocutors minimize the duration of gaps between turns (Kendrick and Torreira, 2014; Sacks et al., 1974), achieving average turn-transition times of 200-300 ms or less (Levinson and Torreira, 2015; Stivers et al., 2009; Dingemanse and Liesenfeld, 2022; Heldner and Edlund, 2010). Qualitative analyses of conversational transcripts have revealed that longer gaps often indicate 'trouble', e.g., foreshadowing disagreement with an assessment, and that they are more likely to occur before dis-preferred responses (e.g., rejections and disagreements) than before preferred responses (e.g., acceptances and agreements) (Pomerantz, 1984; Pomerantz and Heritage, 2012; Stivers et al., 2009). Conversation-analytic work on recordings of naturally occurring conversations that took durational aspects of gaps into account revealed a 'standard maximum' inter-turn silence of approximately one second before interactants treat the gap as troublesome (Jefferson, 1989). Analysis of a corpus of telephone conversations found that dis-preferred responses are more likely than preferred responses after gaps of 700 ms or longer (Kendrick and Torreira, 2014).

These statistical regularities may serve as heuristics in real-time comprehension, i.e., as cues feeding into comprehenders' inferential processes and influencing their judgment of the interaction and the interactants (Templeton et al., 2022; Henetz, 2017; Roberts et al., 2011). For instance, extended gaps lead listeners to doubt respondents' willingness to comply with a request or their agreement with an assessment (Roberts et al., 2006). Systematically varying the duration of gaps between requests and acceptance, Roberts and Francis (2013) identified a "temporal threshold of tolerance" for unmarked inter-turn silence: In their study, gaps exceeding 700 ms triggered inferences about respondents' (un)willingness to comply with the request, i.e., about their stance towards what was requested (A: "Can you pick me up from work?" [silence for 0.2/0.7/1.2] B: "Yeah.")

Extending the scope of gap effects, Henetz (2017) showed that gap durations modulate quali-

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

Speaker A	Speaker B	English translation	
Hi, was geht bei dir?		*Hi, what's up?*	written context
	Hi, ich bin fertig mit dem Training.	*I've finished training.*	
	Kannst du mich abholen?	*Can you pick me up?*	question
(0.20 / 0.72 / 1.20 s)		(0.20 / 0.72 / 1.20 s)	variable silence
Ja.		*Yes.*	positive response

Table 1: Example dialogue in German and its English translation. The contexts were presented visually, whereas the following critical question-response pair was presented auditorily via headphones.

ties that observers of question-response sequences attribute to the questioner, the respondent, and to the interaction as a whole. Specifically, Henetz observed that long inter-turn gaps make the conversation seem more awkward and speakers less engaged, where 'engagement' subsumes speakers' attentiveness, effort, rapport, and interest in future interaction. She proposed that observers iconically map gap durations onto "interpersonal distance" when inferring speakers' stances and affective states (2017, p. 56).

Following up on Henetz's idea of the iconic relation between gap length and interpersonal distance, we report an experiment in German in which we used observers' intuitive judgments to further extend the scope of gap effects to inferences about interpersonal stances, i.e., about "affective stance[s] taken toward another person in a specific interaction, colouring the interpersonal exchange in that situation (*distant*, *cold*, *warm*, *supportive*, *contemptuous*)" (Scherer, 2003). We hypothesized that unexpectedly disfluent turn taking triggers attributional inferences about respondents' personal stance towards the questioner ('the respondent is cold and distant').

Hence, responses preceded by longer gaps should lead participants to judge the respondent as more cold and distant than responses following shorter gaps, in line with earlier reports of gap-based inferencing (Roberts et al., 2006; Henetz, 2017; Fox Tree, 2002; Templeton et al., 2022). If, on the other hand, observers do not rely on inter-turn gap durations as a cue to respondents' interpersonal stance, then distance judgments should be indistinguishable between gap conditions.

We further hypothesized that delay-driven inferences about respondents' interpersonal stance are context-dependent (Whittlesea and Williams, 2001), and will be delayed or attenuated if the situational context provides plausible alternative explanations for the extended inter-turn silence. Specifically, we tested whether comprehenders flexibly adjust their temporal threshold of tolerance to respondents' language proficiency (native vs. non-native speaker), reasoning that observers expect non-native speakers to produce longer inter-turn gaps due to increased difficulty and time requirements for comprehension and response planning, as speech processing is slower in a second language (L2) than in a first language (L1) (e.g. Fitzpatrick and Izura, 2011; Kilborn, 1989). This hypothesis received initial support from a recent study demonstrating an attenuating effect of non-nativeness on observers' judgments of willingness (Matzinger et al., 2023).

We therefore expected that if non-native speakers respond after a longer gap, observers attribute the unexpectedly disfluent floor transfer to these processing constraints rather than interpreting it as evidence of speakers' interpersonal stance. Thus, we predicted an asymmetry between non-native respondents on the one hand, whose long gaps may plausibly be attributed to increased comprehension- and planning time, and native respondents on the other hand, for whom no such 'excuse' was available, so that their delayed responses would more likely be interpreted as evidence of their interpersonal stance. If this assumption is correct, we should observe that the gap effect is delayed (native: short < intermediate/long, non-native: short/intermediate < long) or attenuated (short: native = non-native, intermediate/long: native > non-native) in non-native respondents as compared to native respondents. If, however, the gap effect is unaffected by speakers' command of the language, we should observe indistinguishable gap effects in native vs. non-native respondents. A main effect of nativeness would indicate a general, gap-independent bias.

2 Method

2.1 Materials and design

We constructed 240 short dialogues modeled as the beginnings of telephone conversations between friends. Each dialogue consisted of a brief context followed by a critical question-response pair (= a request/offer and a simple *yes*-response). A German example and its English translation in Table 1 illustrate this template, which we used for constructing all the critical dialogues.

We recorded questions and responses separately and then assembled the question-response pairs, which allowed us to manipulate the duration of inter-turn silence while keeping speech signals identical across gap conditions (Roberts et al., 2006; Henetz, 2017). The contexts were not recorded but presented visually on the screen (see Fig. 1 for an illustration of the trial structure). All recorded speakers were female (four native and four non-native). Questions were recorded only from native speakers of German and response tokens were recorded from both native and non-native (= French) speakers. We selected five response tokens per speaker that were immediately identifiable as (non-)native with a likelihood of > 90% in a pretest ($N_{participants}$ = 44, none of which participated in the main experiment). Note that this pretest-based selection resulted in systematic phonetic differences between native and non-native response tokens, presumably reflecting the cues to non-nativeness used by native listeners (e.g., the duration of individual speech sounds). We therefore statistically controlled for phonetic differences known to affect attributional inferences and the effect that gap durations have on them (see section 2.4 for details).

We varied the gap duration in critical turn transitions (200/720/1200 ms) and the nativeness of the respondent (native/non-native) in a fully crossed 3×2 design. The resulting 1440 dialogues (240 items in six conditions) were distributed across six lists according to a Latin square, so that each participant was presented with each item in only one of the six conditions.

2.2 Participants

We recruited 42 native speakers of German via the recruitment system of the MPI for Empirical Aesthetics, Frankfurt (Germany) and via personal appeal (M_{age} = 27.8, SD_{age} = 6.1, f/m = 22/20). Participants had normal or corrected to normal vision and reported no hearing or reading disabilities. All participants gave written informed consent prior to participation.

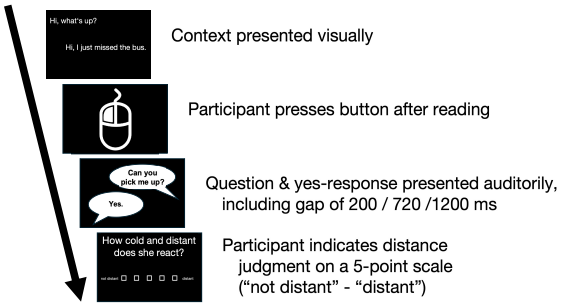

Figure 1: Structure of an experimental trial.

2.3 Procedure

Participants received written instructions on a computer screen and familiarized themselves with their task in two practice trials. Instructions already primed participants that not all of the respondents were native speakers. In each trial, participants first read the context of the dialogue on the screen, which ensured that the critical audio stimulus contained only one turn-timing cue. Once they pressed a button to confirm they had finished reading, the critical question-response pair was presented auditorily via headphones, followed by a visual prompt to intuitively judge "how cold and distant the respondent reacted"[1] and to indicate their judgment on a 5-point rating scale ranging from 'not distant' to 'distant' (Fig. 1). Participants responded by mouse-clicking the appropriate box on the scale and confirmed their definitive response by pressing the space bar on the keyboard. Response latencies were calculated from the onset of the rating task to the mouse click of the confirmed response. Once a response had been confirmed, the next trial started after a blank-screen interval of 1000 ms. The 240 trials were presented in randomized order in three blocks separated by short breaks to reduce participant fatigue (80 trials per block with conditions as equally balanced as possible). The experiment was controlled with *PsychoPy* (Peirce et al., 2019, version 2023.2.3). Experimental sessions lasted about 50 minutes in total. Experimental procedures were approved by the Ethics Council of the German

[1] Original German wording: "Wie kühl und distanziert reagiert sie?"

Linguistic Society (vote 2024-01).

2.4 Analysis

During the experiment, no time limit was set for the judgment task, but we excluded trials with exceedingly long response latencies greater than 10 seconds (< 1% of trials). Response latencies of the remaining trials were log-transformed before the analysis.

Analyses used Bayesian mixed-effects regression to test for fixed main and interaction effects of gap duration (200 ms vs. 720 ms vs. 1200 ms) and respondents' nativeness (native vs. non-native) on distance judgments (ordinal regression) and on response latencies (linear regression); nativeness was sum-coded (native = -0.5; non-native = 0.5); gap duration was treated as a dummy-coded categorical variable with short gap durations (200 ms) as the reference level (intermediate = 720 ms; long = 1200 ms). We specified weakly informative priors for all fixed effects and estimated the models in 10k iterations (ordinal regression) and 6k iterations (linear regression), respectively.

All models contained crossed random effects for participants and items (Baayen et al., 2008). We determined the parsimonious random-effect structure for each model with stepwise forward-fitting, keeping random slopes if evidence for improved model performance exceeded a liberal confidence level of 80% (Bates et al., 2015a); the resulting model of distance judgments contained random slopes for the main effects of gap duration (by-participant) and nativeness (by-item and by-participant). Analyses were conducted in *R* (R Core Team, 2023) using the packages *bayestestR* (Makowski et al., 2019), *brms* (Bürkner, 2017), *emmeans* (Lenth, 2024), *lme4* (Bates et al., 2015b), and *ordinal* (Christensen, 2023).

Note that the response tokens we had preselected to create the native/non-native contrast naturally differed in terms of phonetic parameters (i.e., the phonetic cues of non-nativeness). Specifically, our non-native response tokens had longer durations than the native ones (M_{native} = 264 ms, SD_{native} = 115 ms; $M_{non-native}$ = 477 ms, $SD_{non-native}$ = 109 ms, unpaired *t*-test: $t(38) = 6.01$, two-tailed $p < .001$), and were slightly lower-pitched (M_{native} = 232 Hz, SD_{native} = 30 Hz; $M_{non-native}$ = 214 Hz, $SD_{non-native}$ = 25 Hz, unpaired *t*-test: $t(38) = 2.02$, two-tailed $p = .050$). Earlier results indicated that these parameters affect observers' attributional inferences and modulate effects of extended gap durations, so that they could potentially blur, induce, or cancel out effects of nativeness. Consider, for instance, the observation that longer response durations increase attributions of unwillingness to comply with a request (Roberts et al., 2006); if they also increase attributions of interpersonal distance, then this cue alone could induce an apparent effect of nativeness (greater interpersonal distance for non-native vs. native respondents) while it actually reflects a phonetic cue that equally affects inferences about native and non-native respondents. Such an effect of response duration, in turn, could cancel out or distort the predicted attenuation effect of nativeness (reduced interpersonal distance for non-native vs. native respondents). To disentangle these potentially confounding factors, we tested in a stepwise fashion whether including additional fixed effects of these phonetic parameters improved the model predicting distance judgments, incrementally adding 1) the main effect of response duration, 2) the interaction effect of response duration and gap length, and 3) the main effect of mean pitch, which served as a proxy for the categorical phonetic differences that Roberts et al. had synthesized for their study. We centered pitch values (in 100-Hz bins) and response durations (in 100-ms bins) on the respective median (i.e., model intercepts reflect estimates at their median values and effect estimates reflect duration increases of +100 ms and pitch changes of +100 Hz, respectively). These model comparisons were based on the Bayesian LOO estimate of the expected log pointwise predictive density (ELPD) and took both the observed differences (Δ_{ELPD}) and the standard error of component-wise differences into account (Bürkner and Vuorre, 2019); we kept additional fixed effects if evidence for improved performance exceeded a confidence level of 95%. The resulting model contained an additional fixed main effect of response duration (Δ_{ELPD} = 90.1, 95%-CI$_{\Delta ELPD}$ = [117.4, 62.9]). Bayes factors (BF) estimated in 40k iterations per model provided substantial evidence that the selected model is more likely to have produced the observed data than the base model ($BF = 4 \times 10^{44}$).

3 Results

3.1 Distance judgments

Statistical analysis (see Table 2 for a summary) tested for effects on judgments of interpersonal distance and revealed a main effect of gap dura-

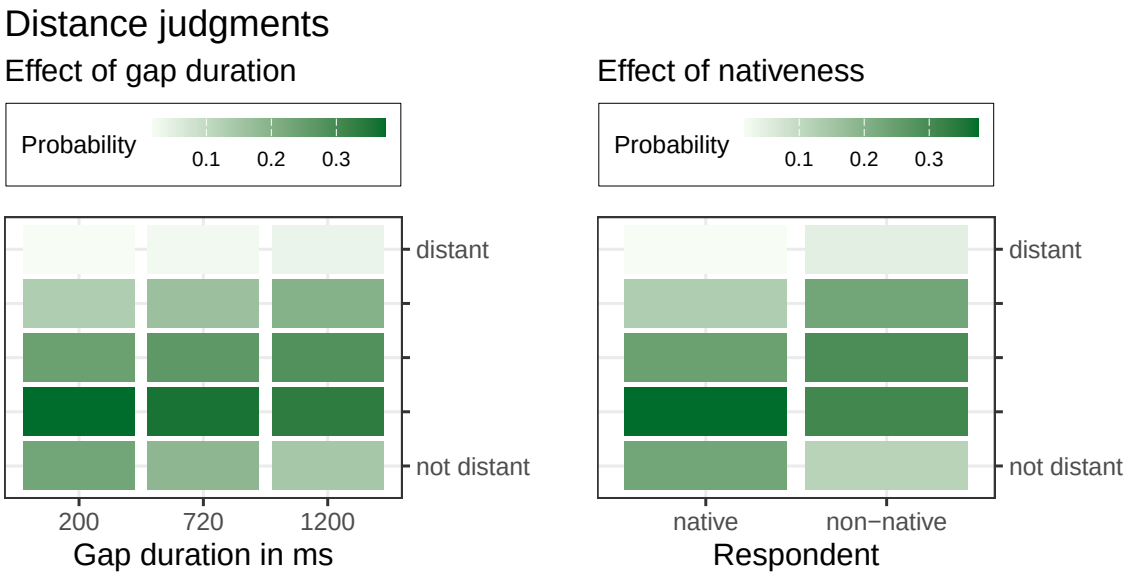

Figure 2: Conditional effects of inter-turn gap duration (left panel) and respondents' nativeness (right panel) on observers' judgments of interpersonal distance; probabilities in each column sum to 1. Attributions of interpersonal distance were more likely for extended inter-turn gaps (200 < 720 < 1200) and for non-native (vs. native) respondents; contrary to our hypothesis, these effects did not interact.

tion: Observers perceived respondents as more cold and distant when inter-turn silence increased from short to intermediate duration (200 ms -> 720 ms: *CE* = 0.16, 95% highest posterior density interval ($HPDI_{95\%}$) = [0.09, 0.23]) and from intermediate to long duration (720 ms -> 1200 ms: *CE* = 0.15, $HPDI_{95\%}$ = [0.07, 0.23]), see Fig. 2, left panel.

Effect	*CE*	*SE*	95% *CI*
Intercept[1]	-0.95	0.10	[-1.15, -0.75]
Intercept[2]	0.06	0.10	[-0.14, 0.27]
Intercept[3]	0.83	0.10	[0.62, 1.03]
Intercept[4]	1.91	0.10	[1.70, 2.11]
gap720	0.16	0.04	[0.09, 0.23]
gap1200	0.31	0.05	[0.20, 0.41]
native	0.46	0.16	[0.15, 0.77]
responseDuration	-0.21	0.04	[-0.28, -0.14]
gap720:native	0.01	0.05	[-0.10, 0.11]
gap1200:native	-0.01	0.05	[-0.12, 0.09]

Table 2: Regression coefficients of Bayesian ordinal mixed-effects regression; *CE* = coefficient estimate, *SE* = standard error, *CI* = credible interval, responseDuration = median-centered duration of the response in 100-ms bins. The model revealed main effects of gap duration, nativeness, and response duration, whereas the estimated slopes of the interaction effect were indistinguishable from zero.

We further observed a main effect of nativeness: Observers perceived non-native speakers' responses as colder and more distant than those of native speakers (*CE* = 0.46, $HPDI_{95\%}$ = [0.15, 0.77], see Fig. 2, right panel). There was no interaction effect, that is, respondents' nativeness did not modulate the effect of inter-turn silence on observers' attributions of interpersonal distance (see Table 2).

Finally, judgments of interpersonal distance were sensitive to the duration of the response token: As shown in Fig. 3, attributions of interpersonal distance were associated with very short response tokens and became less likely when response tokens were longer (CE_{+100ms} = -0.21, $HPDI_{95\%}$ = [-0.28, -0.14]) (see Table 2).

3.2 Response latencies

Linear regression revealed a main effect of gap duration, which reflected that participants responded more slowly when inter-turn gaps exceeded standard expectations (200 ms -> 720 ms: *CE* = 0.11, $HPDI_{95\%}$ = [0.08, 0.14]) and slightly so when inter-turn silence exceeded the intermediate duration (720 ms -> 1200 ms: *CE* = 0.04, $HPDI_{95\%}$ = [0.01, 0.07]); see Figure 4.

There were no main or interaction effects of respondents' nativeness. The results of this linear mixed-effects regression analysis are summarized in Table 3.

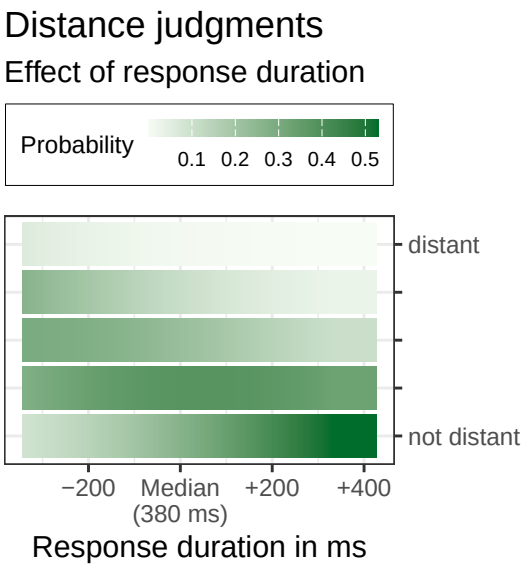

Figure 3: Conditional effect of response duration on distance judgments; durations were centered on the median (380 ms). Short responses increased judgments of interpersonal distance.

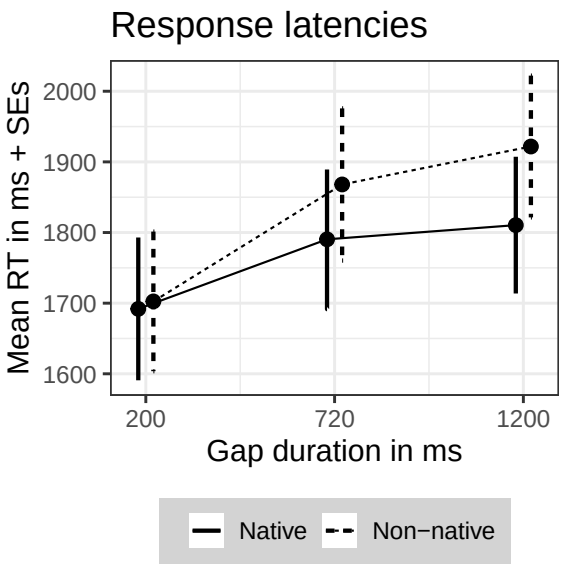

Figure 4: Average response times per condition; error bars indicate standard errors.

Effect	*CE*	*SE*	95% *CI*
Intercept	7.16	0.06	[7.05, 7.27]
gap720	0.11	0.02	[0.08, 0.14]
gap1200	0.15	0.02	[0.12, 0.18]
native	0.04	0.03	[-0.02, 0.09]
gap720:native	0.00	0.03	[-0.06, 0.06]
gap1200:native	0.01	0.03	[-0.05, 0.07]

Table 3: Regression coefficients of Bayesian linear mixed-effects regression predicting log-transformed reaction times; *CE* = coefficient estimate, *SE* = standard error, *CI* = credible interval. Reaction times increased with gap duration but were unaffected by respondents' nativeness.

Discussion

Markedly long turn-transition times in conversation are known to trigger attributional inferences about respondents' stance towards what was put forward, e.g., about their willingnesss to comply with a request or to accept an offer. We tested whether delayed responses in dialogue additionally trigger attributional inferences about interpersonal stances between interlocutors. We had hypothesized that extended inter-turn silence leads observers to attribute to respondents a "cold and distant" interpersonal stance towards the person requesting/offering.

Our results confirm this hypothesis, showing that attributions of interpersonal distance increased from short gap durations (200 ms) to intermediate ones (720 ms), and from intermediate to long ones (1200 ms). While earlier studies had established that gap durations modulate attributions of respondents' stance towards the content of the preceding turn (Roberts et al., 2006) and their evaluation of the interaction (Henetz, 2017), our findings indicate that gap durations also influence attributions of affective interpersonal stances towards the person making the request/offer. Taken together, previous and present results indicate that observers use turn-timing cues for a wide range of attributional inferences, and that delayed responses trigger negative attributions, thus highlighting the importance of smooth turn transitions in conversation. This finding might be relevant for the design and improvement of artificial conversational agents, which partly show slow and invariant turn timing, which could lead to negative attributions by the users of the technology (Barthel, 2025).

Our response-time results also indicate that longer-than-expected inter-turn silence triggers additional mental activity, and thus align with the view that observers use heuristic inferential processes to make sense of turn-taking behavior that violates standard expectations.

We further observed an (unexpected) effect of nativeness, which indicated that observers tended to perceive native respondents as less cold and distant than non-native ones. Aligning with earlier reports of nativeness biases (e.g., Lev-Ari and

Keysar, 2010; Boduch-Grabka and Lev-Ari, 2021), this finding potentially reflects an intra- vs. inter-group effect akin to nativeness effects observed in other types of attributional inferences (Matzinger et al., 2023) and to gap effects on interactants' affective states that differ between racial in-group versus out-group interactions (Pearson et al., 2008). In other words: Our native-speaking participants might have judged those speakers more leniently that belong to the social group they themselves identify with, resulting in the observed bias towards native speakers. Such an in-group heuristic may be restricted to cases where, as in our study, no additional information about the speaker is available beyond the non-nativeness cue. If these post-hoc hypotheses are correct, the nativeness bias should be reversed in a sample of non-native participants performing the same task, and should be overridden by speaker-specific heuristics in an experimental setup that provides participants with sufficient information about specific non-native respondents' levels of language proficiency.

While such a social-identification account seems plausible, the observed penalty for non-native respondents might also reflect (dis)fluency-misattribution (Graf and Landwehr, 2015; Whittlesea, 1993; Whittlesea and Williams, 2001; Reber et al., 2004). In this view, observers process non-native responses less fluently than native ones, resulting in a negative affective response that they misattribute to respondents' interpersonal stance. The indistinguishable reaction times we observed for native and non-native respondents do not support this perspective, but since they do not directly reflect the ease/difficulty of processing the responses they should not be regarded as conclusive evidence. To properly test this (dis)fluency-misattribution account, observers' intuitive judgments would need to be complemented with time-sensitive or physiological measures of processing difficulty.

We had further hypothesized that this gap effect on interpersonal stance is delayed or attenuated if contextual factors provide plausible alternative explanations for a delay in turn transitions. We tested whether respondents' nativeness (L1 vs. L2 speaker) would lead to an asymmetric gap effect, reasoning that non-native respondents might be expected to exhibit longer inter-turn gaps due to additional time needed for comprehension and response planning in a second language (see Matzinger et al., 2023, for evidence in support of this hypothesis).

Though the main effect of nativeness indicates that observers were indeed sensitive to this speaker characteristic, native and non-native speakers' late responses elicited indistinguishable effects of inter-turn silence. Our results thus do not support our hypothesis that the effect of extended inter-turn silence on attributions of interpersonal stance is delayed or attenuated for non-native respondents. Instead, our results suggest that non-native speakers cannot generally count on more lenient evaluations of their turn timing by (native) observers than native speakers.

Recall that our hypothesis was based on the assumption that non-native speakers require more time for comprehension and response planning and thus produce longer inter-turn gaps. This reasoning is called into question by recent evidence indicating that L2 speakers might not necessarily produce longer inter-turn gaps (but take turns less frequently) (Sørensen et al., 2021; Matzinger et al., 2020). L2 speakers rather seem to skip or miss a turn at talk whenever they have not yet planned an appropriate reply by the time their subjective temporal threshold of tolerance has passed, which might cause their interlocutors to self-select for a follow-up turn after an extended gap at positions where turn transition would have been relevant (Davidson, 1984; Sacks et al., 1974). If competent listeners are aware of these regularities, they might not necessarily expect longer gaps in L2 speakers, leading them to apply the same temporal threshold as for native respondents.

Based on the results of earlier investigations, our analysis included phonetic parameters of the response tokens that had been shown to affect observers' attributional judgments. Specifically, Roberts et al. (2006) reported that the pitch contour and the duration of response tokens affect observers' attribution of respondents' (un)willingness to comply with a request. In our study, we also found that the duration of the response tokens affects observers' attributional judgments of interpersonal distance. But whereas Roberts et al. observed that very long response durations decrease the perceived willingness to comply (long duration = negative attribution), we found that very short responses lead observers to infer a cold and distant interpersonal stance (short durations = negative attribution). These apparently contradictory results can be reconciled if we consider the ranges/contrasts of response-token durations used in the two studies. Prioritizing the control of phonetic parame-

ters, Roberts et al. constructed long response tokens by synthetically lengthening natural recordings to twice and triple their duration, which presumably created a hesitation cue that increased the impression of unwillingness. Prioritizing cues to (non-)nativeness, we had used only natural recordings that we pre-selected based on pretest results that estimated how clearly (non-)native they sound. As a result, some of the selected tokens were very short (∼100 ms), which presumably adds a curt, matter-of-fact quality to the response. While this quality was absent in most of our non-native responses, and thus compensates to some degree for the out-group penalty, it apparently conveyed a cold and distant interpersonal stance in very short (mostly native) responses. Hence, we assume that the seemingly inconsistent conceptual mapping of response-token duration onto positive/negative attributes reflects a more general penalty for extreme values: Very short responses convey curtness, very long ones hesitancy.

Conclusion and limitations

We conclude from our rating- and response-time results that extended inter-turn silence triggers negative attributions of interpersonal stance, making respondents seem more cold and distant when they fail to respond timely to a request or an offer. We consider this pattern indicative of inferential processes that are induced by disfluent turn-taking. Moreover, we conclude from observers' intuitive judgments that non-native speakers cannot count on more lenient evaluations of their disfluent turn timing than native speakers. In fact, our results indicate that observers may be biased towards their in-group, attributing greater interpersonal distance to non-native speakers than to native ones. Since phonetic variables partly compensated for this out-group penalty, we believe that additional research is necessary to better dissociate and estimate the relative contributions of these factors, both of which are considered important in the best-performing model identified in our model-selection procedure.

Note, though, that observers had no further information about the actual speakers in our experimental setup, i.e., their judgments were based on a generalized notion of 'non-nativeness'. While this is not an uncommon situation in real life, the effect of nativeness might well be different in interactions in which listeners (= interactants/observers) have more information about *specific* speakers, knowing, for instance, that a particular person frequently struggles to comprehend and to plan and produce appropriate replies in time when speaking the foreign language. While the present results seem to suggest otherwise, it is possible that listeners flexibly adjust their temporal threshold of tolerance even to an individual speaker's command of the language as estimated on the basis of the current or earlier interactions.

Moreover, non-nativeness was represented by only four speakers in our study. While this speaker-sample size mirrors those of related investigations (e.g., Matzinger et al., 2023), larger speaker samples will allow future studies to more confidently generalize their results to other speakers (Clark, 1973).

While the present results failed to corroborate the idea that speakers' linguistic background modulates the effect of delayed turn timing on observers' attributions of interpersonal stance, there could still be other stable speaker characteristics, e.g., gender (Roberts and Norris, 2016), that might have such an effect. Moreover, it seems plausible that transient states of speakers (e.g., momentary inattentiveness) could delay or attenuate such inferences – an empirical question to be addressed in future investigations.

Despite these limitations, our results indicate that observers use response timing and other phonetic parameters as well as speaker characteristics as heuristic cues to respondents' interpersonal affective stance.

Acknowledgments

The research reported in this article has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 497498146. We are grateful to Jürgen Immerz for technical advice and assistance during stimulus recording, to our research assistants and interns Hannah Ida Hullmeine, Theresa Gorski, Karina Valeeva, Kübra Nur Çetin and Gizem Selvi for their help in stimulus construction and data collection, and to the lab team of the MPI for Empirical Aesthetics for participant recruitment support.

References

R. H. Baayen, D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412.

M. Barthel. 2025. Alexa, you are too slow! Invariant turn-transition times and conversational flow in natural human – voice agent interaction. In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue*, Bielefeld, Germany.

D. Bates, R. Kliegl, S. Vasishth, and H. Baayen. 2015a. Parsimonious mixed models. *arXiv:1506.04967 [stat]*.

D. Bates, M. Mächler, B. Bolker, and S. Walker. 2015b. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1):1–48.

K. Boduch-Grabka and S. Lev-Ari. 2021. Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science*, 45(11):e13064.

P.-C. Bürkner. 2017. *brms*: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

P.-C. Bürkner and M. Vuorre. 2019. Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1):77–101.

R. H. B. Christensen. 2023. *ordinal: Regression models for ordinal data*. R package version 2023.12-4.1.

H. H. Clark. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12:335–359.

J. Davidson. 1984. Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In J.M. Atkinson and J. Heritage, editors, *Structures of social action: Studies in conversation analysis*, pages 102–128. Cambridge University Press, Cambridge.

M. Dingemanse and A. Liesenfeld. 2022. From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5614–5633, Dublin, Ireland. Association for Computational Linguistics.

T. Fitzpatrick and C. Izura. 2011. Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition*, 33(3):373–398.

J. E. Fox Tree. 2002. Interpreting pauses and ums at turn exchanges. *Discourse Processes*, 34(1):37–55.

L. K. M. Graf and J. R. Landwehr. 2015. A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19(4):395–410.

M. Heldner and J. Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

T. Henetz. 2017. *Don't hesitate! The length of inter-turn gaps influences observers' interactional attributions*. PhD Thesis, Stanford University, Stanford.

G. Jefferson. 1989. Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger and P. Bull, editors, *Conversation: An interdisciplinary perspective*, pages 166–196. Multilingual Matters, Philadelphia.

K. H. Kendrick and F. Torreira. 2014. The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4):1–35.

K. Kilborn. 1989. Sentence processing in a second language: The timing of transfer. *Language and Speech*, 32(1):1–23.

R. V. Lenth. 2024. *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.10.5.

S. Lev-Ari and B. Keysar. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6):1093–1096.

S. C. Levinson and F. Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6(731):10–26.

D. Makowski, M. S. Ben-Shachar, and D. Lüdecke. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, 4(40):1541.

T. Matzinger, M. Pleyer, and P. Żywiczyński. 2023. Pause length and differences in cognitive state attribution in native and non-native speakers. *Languages*, 8(1).

T. Matzinger, N. Ritt, and W. T. Fitch. 2020. Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. *PLOS ONE*, 15(4):e0230710.

A. R. Pearson, T. V. West, J. F. Dovidio, S. R. Powers, R. Buck, and R. Henning. 2008. The fragility of intergroup relations: Divergent effects of delayed audiovisual feedback in intergroup and intragroup interaction. *Psychological Science*, 19(12):1272–1279.

J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203.

A. Pomerantz. 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J.M. Atkinson and J. Heritage, editors, *Structures of Social Action: Studies in Conversation Analysis*, pages 53–101. Cambridge University Press, Cambridge.

A. Pomerantz and J. Heritage. 2012. Preference. In J. Sidnell and T. Stivers, editors, *The Handbook of Conversation Analysis*, pages 210–228. John Wiley & Sons, Ltd, Chichester, UK.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

R. Reber, N. Schwarz, and P. Winkielman. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4):364–82.

F. Roberts and A. L. Francis. 2013. Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, 133(6):EL471–EL477.

F. Roberts, A. L. Francis, and M. Morgan. 2006. The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, 48(9):1079–1093.

F. Roberts, P. Margutti, and S. Takano. 2011. Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, 48(5):331–354.

F. Roberts and A. Norris. 2016. Gendered expectations for "agreeableness" in response to requests and opinions. *Communication Research Reports*, 33(1):16–23.

H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

K. R Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1):227–256.

T. Stivers, N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. de Ruiter, K. E. Yoon, and S. C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10587–10592.

A. J. M. Sørensen, M. Fereczkowski, and E. N. MacDonald. 2021. Effects of noise and second language on conversational dynamics in task dialogue. *Trends in Hearing*, 25:23312165211024482.

E. M. Templeton, L. J. Chang, E. A. Reynolds, M. D. Cone LeBeaumont, and T. Wheatley. 2022. Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4):e2116915119.

B. W. A. Whittlesea. 1993. Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6):1235–1253.

B. W. A. Whittlesea and L. D. Williams. 2001. The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1):3–13.