

# Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer

WEN DUAN, Cornell University, USA & NTT, Japan

NAOMI YAMASHITA, NTT, Japan

SUSAN R. FUSSELL, Cornell University, USA

Collaborating<sup>1</sup> using a common language can be challenging for non-native speakers (NNS). These challenges can be reduced when native speakers (NS) adjust their speech behavior for NNS, for example by speaking more slowly. In this study, we examined whether the use of real-time speech rate feedback (a speech speedometer) would help NS monitor their speaking speed and adjust for NNS accordingly. We conducted a laboratory experiment with 20 triads of 2 NS and 1 NNS. NS in half of the groups were given the speech speedometer. We found that NS with the speech speedometer were significantly more motivated to slow down their speech but they did not actually speak more slowly, although they made other speech adjustments. Furthermore, NNS perceived the speech of NS with the speedometer less clear, and they felt less accommodated. The results highlight the need for tools that create scaffolding to help NS make speech accommodations. We conclude with some design ideas for these scaffolding tools.

CCS Concepts: **Human-centered computing** → **Collaborative and social computing** → **Empirical studies in collaborative and social computing**

**KEYWORDS:** Computer-mediated Communication; Multilingual Communication; Cross-lingual Communication; Speech Accommodation

## ACM Reference format:

Wen Duan, Naomi Yamashita and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3, No. CSCW, Article 171 November 2019. ACM, New York, NY, USA. 25 pages. <https://doi.org/10.1145/3359273>.

## 1 INTRODUCTION

Global teams often use a common language such as English to communicate, and this can create burdens for nonnative speakers (NNS) of the common language because using a foreign language can decrease their ability to think and process information [61]. NNS whose professional knowledge and skills are valuable to the collaboration may not be able to contribute as much as they could in their native language [37][69] due to difficulties in following, understanding and participating in the often fast-paced team conversations in the common language. Especially when

Authors' email addresses: [wd238@cornell.edu](mailto:wd238@cornell.edu), [naomiy@acm.org](mailto:naomiy@acm.org), [sfussell@cornell.edu](mailto:sfussell@cornell.edu)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

2573-0142/2019/November - ART171 \$15.00

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

<https://doi.org/10.1145/3359273>

native speakers (NS) are predominant in number, they tend to carry the conversation forward among themselves at a speed faster than they would speak with a NNS one on one [6][20], particularly when the discussion is heated [45][69].

Research shows that NS are motivated to accommodate by making their speech clearer for NNS, for example by reducing their speech rate [2][18][32][52]. However, vocal characteristics like speech rate are often outside of speakers' conscious control, especially when they are excited or trying to be persuasive [1][49]. In addition, even when NS consciously attempt to accommodate, their speech modifications may not be appropriate or effective [60] because NS don't know what level of accommodation is needed, or are not aware of how quickly they are speaking in comparison to what is needed.

While significant previous work has explored tools to aid NNS in multiparty conversations [18][19][32][52][65], less work has aimed to facilitate NS's efforts to accommodate for NNS team members (e.g., by helping NS understand the best ways to accommodate). In this paper, we focus on supporting NS's efforts to accommodate NNS in multiparty communication. We propose a tool, the Speech Speedometer, that provides real-time evaluative feedback on NS speech behavior. The tool aims to facilitate one type of accommodation, slowing down speech, by showing NS their real-time speech rate using a bar speedometer that indicates slow speech in green and fast speech in red.

Using a laboratory study, we examined the effects of incorporating our Speech Speedometer in a videoconferencing interface on speech rate and subjective ratings of conversation experience. Twenty triads of 2 NS and 1 NNS performed three collaborative decision-making tasks, counterbalanced across three trials. NS participants in half of the triads were given our Speech Speedometer in the second trial; those in the control group were not. We found that NS with the speech speedometer were significantly more motivated to slow down their speech. However, they did not actually speak more slowly, although they made other speech adjustments such as using significantly shorter utterances. This was true both during the trial when the tool was present and in the subsequent trial when it was removed. We discuss how this speech adjustment affected NNS subjective evaluations of their communication experience. We also discuss the implications of our findings for the design of tools to support multilingual conversation.

## 2 RELATED WORK

In this section, we first review previous work on the challenges faced by both NNS and NS in multiparty communication. Then, we describe previous systems that were designed to elicit change in speech behavior through real-time feedback. Finally, we review findings in cross-lingual and accommodation research that formed the bases of the mechanisms and parameters of our tool to help NS monitor and adjust their speech rate in conversation with NNS.

### 2.1 Challenges in Multiparty Communication between NS and NNS

Although using a common language enables people who don't speak the same native language to communicate and collaborate across national and cultural boundaries, studies have shown that issues of language proficiency disadvantage NNS in group communication involving both NS and NNS, and decrease group communication efficiency [47][69][71]. Using a foreign language increases NNS's cognitive load because they need to spend extra effort interpreting and formulating messages [61]. As a result, NNS often miss their chance to express their thoughts in the stream of speech [56]. In fact, Yamashita and colleagues [69] found that the majority of

Japanese scholars who serve as international committee members in their field reported that they cannot even contribute half of what they want to express in English in global meetings.

The burden of using a nonnative language not only affects NNS, it creates problems for the entire team. NS team members are often unaware of the language difficulties faced by NNS and how profoundly they can affect NNS contributions to a conversation, leading to erroneous conclusions about NNS attitudes, beliefs, and personality [32]. This lack of awareness is exacerbated when NS outnumber NNS in a group conversation, because the fluid conversation among NS can create the illusion that conversation flows well among all members [32][46][69]. Further, for reasons such as maintaining face, NNS are reluctant to interrupt NS for clarification or signal them to slow down [56] and they sometimes even send confusing signals that lead NS to think they understand when they do not [68]. Relying on NNS reactions could thus lead NS to think their NNS partners need no accommodation.

Significant previous work has focused on exploring tools to aid NNS in multiparty collaborations. The most widely adopted approach is to provide NNS with automated transcripts, which improves comprehension in audio and video meetings [53], and helps NNS catch up on missed conversation [31]. These approaches have both benefits and costs for NNS, because reading automated transcripts while following the ongoing conversation can create greater cognitive load. To allow NNS more time to process the fast paced conversation with NS, Yamashita et al. [69] created artificial silence gaps in real-time conversation, which resulted in better understanding and higher participation for NNS but also had detrimental effect such as creating speech overlaps. Surprisingly few studies have looked at and made use of NS' willingness to accommodate for NNS in multilingual communication. This is unfortunate because NS may have much more cognitive resources available to allocate to additional tasks while communicating in their native language.

Only a handful of studies have considered the role of NS in designing tools to support NNS. Duan and colleagues [13] employed an intelligent agent that asked clarification questions so that NS were aware of potential confusions that NNS may be too shy to acknowledge. Gao et al. [19] provided NS with a shared display of how their NNS teammates were using real-time transcripts and a bilingual dictionary and found that this increased NS's awareness of the specific difficulties NNS were experiencing. Still, these tools engaged NS only at the awareness level. To our knowledge, one of the few approaches that engaged NS actual effort was that of Pan et al. [52], who found that NS were motivated to highlight keywords in automated transcripts for NNS, thereby balancing the workload for NS and NNS. So far, there is lack of research on how to stimulate and utilize NS initiative and motivation to help NNS in multiparty conversation by making speech accommodations, let alone a tool to facilitate this process. Even when NS are aware that NNS need them to adapt their speech behavior, they often do not know the best way to do so [60]. For instance, how slowly should they speak in order for NNS to understand? Does this differ by topic or specific conversational partner? In addition, NS are often unaware of their own speech behavior, for instance whether they are speaking at a fast, average or slower rate than usual, as speech rate is sometimes out of a person's conscious control especially when he or she is stressed, nervous, excited or trying to be persuasive [1][45][49].

## 2.2 The Use of Real-Time Feedback

Moment-by-moment feedback can help people dynamically adjust their behavior as needed [39]. In the context of multilingual conversation, real-time feedback might both make NS more aware

of their vocal behaviors and give them immediate confirmation as to whether attempts to modify these behaviors are successful [16]. Previous work has explored ways to provide real-time feedback on speech behavior for achieving ideal (desirable) performance both at individual and at group level. For instance, Tanveer et al. [62] developed a wearable tool that helps speakers moderate their speech rate and vary their volume to improve public speaking performance. Through iterative brainstorming, crowdsourcing and testing, their final design succeeded in providing feedback on two speech dimensions (speech rate and volume) during live speech without increasing cognitive load for the user. Hoque et al. [36] designed a system that provides real-time feedback on speech, prosody and nonverbal behaviors through a virtual agent and visualization of the user's performance on job interview practice. DiMicco et al. [12] found that a shared visual display of individual participation rate influenced the amount of time speaking for over and under-participating group members. Leshed et al.'s [44] *GroupMeter*, a system that provided real-time visualizations of word count and agreement words, led to greater reflection on language use and more consistency within teams. To our knowledge, only Ye et al. [70] has proposed a system that was designed with the purpose of raising native speakers' awareness of their speech speed, but they did not test this system with users.

### 2.3 Real-Time Speech Rate Feedback Using the Speech Speedometer

Based on this previous work, we supposed that increasing NS's awareness of their speech behavior and providing real-time evaluative feedback on that behavior were key to motivating and helping NS make speech adjustments. We developed an interface that provides NS with both an awareness of a recommended adjustment strategy and an evaluation of the speaker's own speech behavior in comparison to the target behavior. Specifically, our tool displays the real-time speech rate of the speaker against a recommended speech rate (see Figure 1 and 3.4.2 for details).

Our tool focused on speech rate among other speech dimensions because prior work suggests that a lower speech rate increases NNS comprehension (e.g., [11][57]). For instance, studies such as [5][27][41][58] found that listeners regardless of language background found slow speech to be higher in clarity than fast speech. Griffiths [26] found that moderately fast speech rates over 150 wpm reduced NNS comprehension significantly but comprehension scores did not differ significantly for passages delivered at a slow speech rate (100 wpm) vs. speech rates between 100 and 150 wpm. Other researchers (e.g., [39]) report that speech at 124 wpm improves intermediate level speakers' abilities significantly. Despite the lack of agreement on a specific threshold rate for improving NNS comprehension, prior work consistently shows that slowing down speech will be beneficial for NNS. In addition, the speech accommodation literature suggests that listeners prefer speech rates that are similar to their own speaking rate [7]. We therefore empirically determined a speech rate parameter at 140 wpm for our evaluative Speech Speedometer based on a pilot study. In the pilot study, we calculated the average speech rate of 3 Japanese who spoke English as second language doing the same task as in the main study.

We hypothesized that the real-time evaluative speech rate feedback (Speech Speedometer) would be effective in two ways:

**H1:** NS with the Speech Speedometer will feel more motivated to slow their speech rate for NNS than NS without Speech Speedometer.

**H2:** NS with the Speech Speedometer interface will speak more slowly than NS without interface.

Previous studies [5][11][25] have also found that NS adjustment of speech rate is a function of articulation rate, number and length of pauses. In doing a narration task to NNS [11], NS were

found to increase their total amount of pause time. In our study, we handled pauses of different lengths in defining our speech rate measures such that pauses shorter than 1.2s are reflected in the speech rate (e.g. speech rate will decrease as speakers add more pauses shorter than 1.2s), and that pauses longer than 1.2s will result in shorter utterance length (see 3.4.1 for details). We are interested in exploring:

**RQ1:** How will the Speech Speedometer interface affect NS pausing behavior? Will they add more short pauses (pauses shorter than 1.2s, which will be reflected in the testing of H2)? Or will they add more long pauses (pauses of or longer than 1.2s), which will result in shorter utterances?

We intended for our design of the Speech Speedometer to reduce NSs' cognitive load by providing feedback on only one information dimension (as compared to two dimensions in Tanveer et al.[62]'s design). However, in Tanveer et al.'s study, users performed a one-way speaking task. It is thus unclear whether and how the interface will affect NS users' cognitive load during an interactive conversation. We therefore ask:

**RQ2:** Will the Speech Speedometer affect NSs' workload and communication effort?

In addition to the effects on NS's speech, we anticipate potential effects of the Speech Speedometer on NNS, too. It is worth noting that previous studies investigating effects of speech rate on NNS comprehension conducted their studies using one-way listening tasks [4][7][11][26][39]. We based our design of the interface and hypotheses upon these findings, since communication through videoconferencing requires listening to begin with. Along the lines of research on speech rate and perceived clarity of speech (e.g., [5][41]), we hypothesize that:

**H3a:** NS speech rate will be negatively associated with NNS perceived clarity of NS speech.

**H3b:** NNS perceived clarity of NS speech will be higher in the Speech Speedometer condition than in the control condition.

We extend the concept of listening comprehension in one-way listening tasks to NNS understanding of multiparty discussion. We hypothesize that:

**H4a:** NS speech rate will be negatively associated with NNS understanding of multiparty discussion.

**H4b:** NNS will have better understanding of the multiparty discussion in the Speech Speedometer condition than in the control condition.

Unlike reading or lecture speech, conversational speech ranks the fastest among different types of daily speech [63], which could make NNS all the more cognitively laden. Slowing down can provide NNS more time to process what is being said, which in turn may reduce their cognitive workload and comprehension effort [69]. In a conversational setting, we broaden the notion of comprehension effort to "communication effort", which includes the effort to follow, understand and contribute to an ongoing conversation. We then hypothesize:

**H5a:** NS speech rate will be positively associated with NNS workload and communication effort.

**H5b:** NNS will report lower *workload* and *communication effort* in the Speech Speedometer condition than in the control condition.

Finally, research studying the perception of speech accommodation ([7][21][23][59]) have found that receivers (listeners) recognizing speakers' speech rate convergence perceive the speaker and his/her adjustment more favorably than those recognizing divergence. Needless to say, NS speech rates are generally much faster than that of NNS, creating huge divergence. We therefore hypothesize:

**H6a:** NS speech rate will be negatively associated with NNS perceived NS accommodation.

**H6b:** NNS will feel more *accommodated* by NS in the Speech Speedometer condition than in the control condition.

### 3 METHODS

#### 3.1 Experiment Design

We tested our hypotheses in a laboratory experiment using a 2 (condition: with or without the Speech Speedometer) by 3 (trial) design. Twenty triads consisting of 2 NS and 1 NNS participated in a series of legislative dilemma tasks using a three-way video conferencing tool. Participants were randomly assigned to either the treatment group, where NS speakers were given Speech Speedometer in the second trial, or the control group, where NS were given an equivalent stimulus (using pre-recorded dummy data from the treatment group to ensure the same level of distraction) in the second trial and told that the display reflected network traffic exchange. The tasks and experiment conditions were fully counterbalanced to avoid ordering effects.

#### 3.2 Participants

A total of 60 participants were recruited through a third-party personnel company. Among them 40 (12 female) were native English speakers who grew up and received their education in an English-speaking country (e.g. USA, UK, Australia, etc.). They had lived in Japan for 0.5 to 29 years ( $M=13.58$ ,  $SD=8.24$ ) and reported having very frequent communication with nonnative speakers of English in their daily life ( $M=6.20$ ,  $SD=1.44$  on a scale from 1=never to 7=very often). The mean age for NS was 45.37 ( $SD=11.73$ ).

The remaining 20 participants (15 female) were native speakers of Japanese who spoke English as a second language. All had lived in an English-speaking country for less than two years and rarely communicated with native English speakers in their daily life ( $M=2.45$ ,  $SD=1.23$  on a scale from 1=never to 7=very often). We required that their TOEIC score fall between 700-800, which is considered intermediate to advanced level English ability. Their mean age was 35.65 ( $SD=15.06$ ).

#### 3.3 Materials

*3.3.1 Task.* Our primary criterion for choosing the task was that it would get NS to talk at their natural speed and fully engage in the task, which simulates many multiparty multilingual collaborations [14]. Tasks that involved activities other than talking such as describing tangrams, map navigating, etc. would not do, since multitasking slows down speech rate [64]. Our second criterion was that the task must allow all participants to discuss freely with no assigned roles or prescribed order of speaking.

Based on these criteria, we chose a Legislative Dilemma Task [50] and adapted it into three sets of tasks on three different issues for discussion. Participants were told to imagine that they were UNESCO (United Nations Educational, Scientific and Cultural Organization) representatives who had \$1.8 billion funding to allocate among five competing projects. The funding constraints only allow for supporting two projects, with one of them (their first choice) fully funded at \$1 billion and the other (their second choice) 80% funded at \$0.8 billion. For each trial, participants chose from five candidate projects on a different issue – culture, health and education. They were asked to first choose two projects on their own, then communicate their individual decision with the group to convince them that their proposal was the best, and finally come to an agreement as a group.

Projects were adapted from the UNESCO website after pilot testing. In order to foster discussion, each set of projects included one project likely to cause disagreement between NS and Japanese NNS ([48][50]). None of the projects were located in Japan or countries where NS participants were likely to come from in order to reduce potential bias. Sample project candidates

include: strengthening minority language publishing in East Africa (culture), helping youth cope with mental health issues arising from traumatic experience of war and humanitarian emergencies (health), and supporting the development of female technical talents in Latin America (education).

**3.3.2 Survey.** Participants completed four surveys during the study: a *pre-experiment survey*, in which we collected demographic information, frequency of communicating with NS/NNS, etc. before the first trial; a *post-task survey* that included questions about participants' workload, communication effort, etc. after each trial. All survey data was collected using an online survey tool. Survey questions were shown in English with Japanese translation.

**3.3.3 Interview.** We developed an open-ended interview protocol for participants which asked them to reflect on their communication experience. For NS, we asked additional questions regarding their speech adjustment and reaction to the Speech Speedometer.

## 3.4 Software and Equipment

**3.4.1 Speech recognition tool and speech rate calculation.** We used an ASR (automatic speech recognition) tool developed by NTT for calculating speech rate. An "utterance" is defined by the tool as a speech segment separated by an acoustic silence. The ASR technology detects audio input for the start and end of an utterance. The end of the utterance is detected when the speech signal drops below a certain frequency threshold (acoustic silence) for a specified duration. Many end-of-utterance detectors in speech technology use an acoustic *silence threshold* of 500ms by default so that speech recognition is updated frequently [33]. However, in order to provide meaningful feedback to NS, we adapted the silence threshold to 1.2s to ensure the system captured entire speech turns. Note that there was a tradeoff between providing meaningful feedback and real-time feedback.

We monitored each speaker's speech separately and calculated the speech rate for each utterance as: the number of transcribed words divided by the time duration the utterance was delivered. Although syllables per minute is more accurate in calculating speech rate, most ASR technology does not support syllable level transcription. The adjustment of the silence threshold from 0.5s to 1.2s serves another function to account for pauses. This adjustment allows the calculation of the speech rate of each utterance to account for pauses shorter than 1.2s. The speech rate will decrease as the number of pauses shorter than 1.2s increases. However, pauses longer than 1.2s will not be reflected in the speech rate, but rather result in splitting the speech stream into two separate utterances. We also excluded utterances that were less than 3 words, because those were likely noise transcribed into texts, or backchannel messages such as "um hmm" or "I see", the delivery of which is often very fast but does not affect NNS comprehension.

**3.4.2 Videoconferencing Interface with Speech Speedometer.** Drawing on Tanveer et al. [62], we designed our real-time Speech Speedometer interface as a bar speedometer centered on each NS's screen, with green bars suggesting that he or she was successfully maintaining a slow speech rate and red bars indicating that he or she was going too fast and needed to slow down (see [Figure 1](#)).

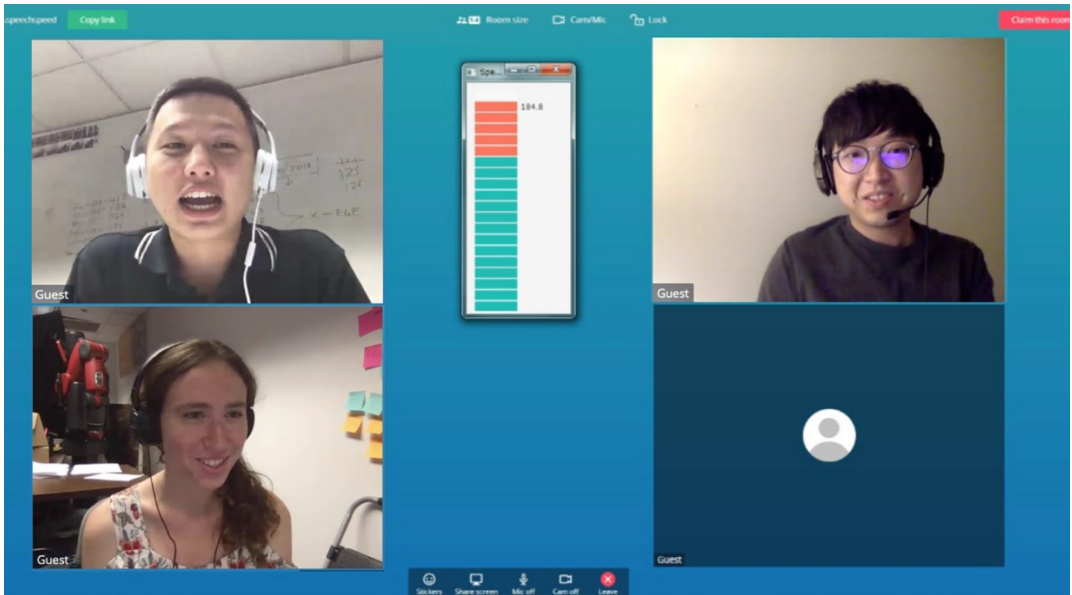


Figure 1: Videoconferencing interface with real-time evaluative Speech Speedometer. NS see their own speech rate shown on the bar in the center of their screen. It updates every time they make a pause longer than 1.2s. The height of the bar suggests the speed of the utterance prior to the pause. The actual speech rate is shown next to the top of the bar. The color provides evaluative feedback on whether the speech rate is too fast or acceptable. When the speech rate exceeds 140 wpm, the bars above the threshold become red. When the NS is not speaking, the bar stays at the speech rate of their last utterance and does not reset.

**3.4.3 Equipment setup.** We video- and audio-recorded each participant's screen activities, including their verbal and nonverbal behaviors over the video-mediated conversations. Three 3D cameras were located behind each participant to video and audio record the entire session including interviews.

### 3.5 Procedure

Each session lasted about 2.5 hours, during which participants were given instructions, completed a preliminary survey, did 3 collaborative decision-making tasks over a three-way videoconferencing platform, completed a post-task survey following each task, and participated in a face-to-face semi-structured interview. The Speech Speedometer interface and its equivalent control display was present only in Trial 2 (see [Figure 2](#)).

Participants arrived in triads (2 NS and 1 NNS) and were led to separate soundproof rooms. General introductions and instruction were given in English over audio by the first author. NNS were assisted by a native Japanese-speaking experimenter who gave clarifications in Japanese.

After signing the informed consent, participants completed a demographic survey. They then started to work on making their individual choices for the first task. Once all of them made their individual choices, they started the free discussion. The experimenter interrupted the conversation after 15 minutes, at which point participants moved on to the post-task survey.

The first trial was used for participants to get familiar with each other and the task. The Speech Speedometer or the equivalent control display was introduced in the second trial. NS in both conditions were told that speaking slowly could help nonnative speakers follow and understand better. NS in the Speech Speedometer condition were also told that they would see a speech speedometer in the center of their screen that showed whether they were talking too fast.



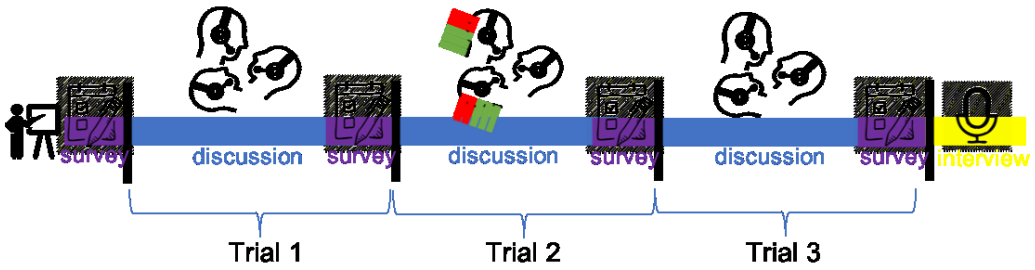


Figure 2: Study procedure. After general instruction, participants started with a preliminary survey. Then they worked on the task individually followed by a 15-minute discussion to come to a group decision. After the discussion, they completed a post-task survey. This task-discussion-survey repeated for 3 trials. In the 2<sup>nd</sup> trial, NS were given the Speech Speedometer interface or the equivalent control display. Upon completion of the last survey, participants were interviewed individually.

If their speech rate exceeded 140 words per minute, which is a recommended rate for talking to NNS, the speedometer bar would turn red. If it's no faster than the threshold, the bar would stay green. NS were told that they could choose whether or not to adjust their speech rate. The Speech Speedometer or control interface was removed for the third trial so we could assess carryover effects. NNS were unaware of the feature throughout the experiment and NS did not know if others in their group had the display or not.

After completion of all three discussions and post-task surveys, we conducted semi-structured interviews in participants' native language asking about their experience.

### 3.6 Measures

We collected both objective data about speech rate and message length and subjective survey measures after each trial.

**3.6.1 NS motivation to slow for NNS.** Starting from the end of Trial 2, NS responded to the statement "I felt motivated to speak slowly when talking to the nonnative speaker" on a 7-point Likert scale (1=strongly disagree, 7=strongly agree). We did not present this measure in Trial 1, to prevent NS from being primed to slow down in the subsequent trial where we implemented the Speech Speedometer or its equivalent display.

**3.6.2 NS speech rate.** The speech rate for each NS for each session was calculated as the total number of words spoken over the total duration of time speaking. Note that speech recognition and speech-rate calculation systems were activated for every discussion even when the speedometer was not shown to the participants.

**3.6.3 NS utterance length.** Utterance length was calculated by averaging the number of words per utterance a NS spoke in each trial.

**3.6.4 Workload.** After each trial, participants completed the NASA-TLX [28], a widely used tool to measure subjective task workload, using a 7-point Likert scale (1=extremely low, 7=extremely high). We dropped the "physical" dimension out of the 6 dimensions (mental, physical and temporal demand, performance, effort and frustration) since our task didn't involve physical activity. The questions formed a reliable scale (Cronbach's  $\alpha = .78$ ) and were averaged to form a workload measure.

**3.6.5 Communication effort.** We constructed a three-item scale to measure communication effort. Participants rated the amount of effort required to "follow the conversation" "fully engage in the conversation" and "contribute to the conversation" on 7-point (1=extremely low,

7=extremely high) Likert scales. They formed a reliable scale (Cronbach's  $\alpha = .97$ ) and were averaged to create our measure of communication effort.

**3.6.6 NNS perceived clarity of NS speech.** This was measured using one item. "Person A/B spoke clearly enough for me to understand" and NNS rated on a 7-point (1=strongly disagree, 7=strongly agree) Likert scale. There were two outliers in the distribution, which were truncated to the mean  $-2.5SD$ , following recommended data transformation procedures (e.g., [9]).

**3.6.7 NNS understanding of multiparty discussion.** We asked participants to rate on 7-point (1=strongly disagree, 7=strongly agree) Likert scales for the following 3 statements: "I could follow my partners' ideas easily", "I had difficulty understanding my partners' opinions" (reversely coded), "I could understand each partner's thoughts well". They formed a reliable scale (Cronbach's  $\alpha = .87$ ) and were averaged to create our measure of NNS understanding of multiparty discussion.

**3.6.8 NNS perceived NS accommodation.** We measured NNS's perception of each NS's accommodation using a 4-item scale adapted from Gasiorek [20]. NNS responded on a 7-point Likert scale (1=strongly disagree, 7=strongly agree) about whether they felt each NS "adjusted his/her ways to communicate for me", "adjusted his/her communication appropriately", "was intending to be helpful" and "was having good intention". We truncated two low outliers to the mean  $-2.5SD$ . The questions formed a reliable scale (Cronbach's  $\alpha = .89$ ) and were averaged to form our measure of perceived accommodation.

**3.6.9 Manipulation check of NS awareness of the Speech Speedometer.** NS awareness of the Speech Speedometer was assessed by a multiple choice question asking them to identify what was displayed on their screen during the second discussion ("I was aware of the real-time speech rate feedback" "I was aware of the network data traffic" "I wasn't aware of any of the above").

## 4 RESULTS

To test our hypotheses and answer our research questions, we performed a series of 2 (condition: presence vs. absence of Speech Speedometer) by 2 (Trial 2 vs. Trial 3) Mixed Model ANOVAs. Participants were nested within groups as a random factor, and the presence of Speech Speedometer and trial were set as independent fixed variables. Trial 1 served as a warmup before the introduction of the Speech Speedometer so it was not included in the analyses.

### 4.1 Manipulation Check

A manipulation check of NS awareness of whether they had received the Speech Speedometer vs. the control display showed that our manipulation was successful. All NS (100%) in the treatment condition correctly identified the display shown on their screen as "real-time speech rate feedback" and all in the control condition correctly identified the display as "network data traffic".

### 4.2 Effects of Speech Speedometer on NS Speech Rate and Pausing Behavior

H1 hypothesized that NS would feel more motivated to slow down their speech with the Speech Speedometer than with the control display. To test this, we conducted a 2x2 Mixed Models ANOVA on *NS motivation to slow for NNS*. Consistent with H1, there was a significant main effect of Speech Speedometer ( $F [1, 38] = 3.86, p = .05$ ), suggesting that NS felt more motivated to slow down their speech for NNS with the Speech Speedometer ( $M=5.05, SE=.38$ ) than without ( $M=4.1, SE=.35$ ). There was also a significant main effect of trial ( $F [1, 38] = 7.92, p < .01$ ), indicating that NS felt more motivated to slow down in Trial 3 ( $M=4.83, SE=.35$ ) where the Speech Speedometer and its equivalent control display were removed, than in Trial 2 ( $M=4.33, SE=.38$ ) where the

displays were present. This was further qualified by a significant interaction effect ( $F [1, 38] = 7.92$ ,  $p < .01$ ) between condition and trial (see [Figure 3](#)). NS in the control condition felt significantly less motivated to slow down for NNS in Trial 2 ( $M=3.6$ ,  $SE=.36$ ) than in Trial 3 ( $M=4.6$ ,  $SE=.33$ ), whereas there was no effect of trial for NS with the Speedometer display (T2:  $M=5.05$ ,  $SE=.39$ ; T3:  $M=5.05$ ,  $SE=.37$ ) where they felt equally motivated in both trials.

H2 hypothesized that NS provided with the Speech Speedometer would speak more slowly than NS without the feedback. To test this hypothesis, we conducted a 2 by 2 Mixed Models ANOVA on NS speech rate. There was no significant main effect of Speech Speedometer ( $F [1, 38] = .30$ ,  $p = .59$ ), no main effect of trial ( $F [1, 38] = .86$ ,  $p = .36$ ) and no interaction ( $F [1, 38] = 2.08$ ,  $p = .16$ ). Contrary to H2, NS did not speak significantly more slowly when given the speedometer.

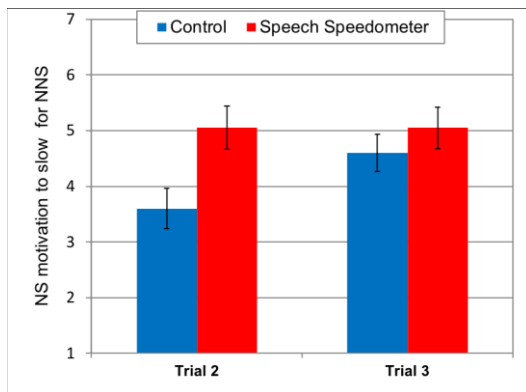


Figure 3: Mean NS Motivation to Slow for NNS on a scale of 1 to 7, by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean)

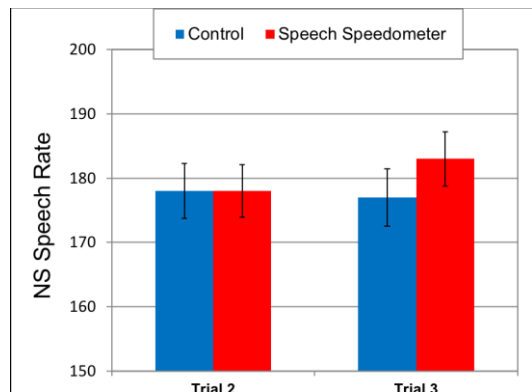


Figure 4: Mean NS Speech Rate, by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean).

To explore RQ1, which asked about the effect of the Speech Speedometer on NS pausing behavior, we conducted a 2 by 2 Mixed Models ANOVA on NS utterance length. We found a significant main effect of Speech Speedometer ( $F [1, 38] = 7.68$ ,  $p=.01$ ), no main effect of trial ( $F [1, 38] = 1.92$ ,  $p=.17$ ) and a trend toward a significant interaction between Speech Speedometer and trial ( $F [1, 38] = 3.01$ ,  $p = .09$ ). NS who were given the Speech Speedometer produced significantly shorter utterances ( $M=9.66$ ,  $SE=.64$ ; see [Figure 5](#)) than those who were not given the Speech Speedometer ( $M=12.14$ ,  $SE=.75$ ). NS with the Speech Speedometer produced significantly shorter utterances in Trial 2 ( $M=9.02$ ,  $SE=.59$ ) than in Trial 3 ( $M=10.31$ ,  $SD=.69$ ). Overall, the results suggested that NS with the Speech Speedometer produced shorter utterance, but their speech rate within the utterance was not significantly slower than the control group.

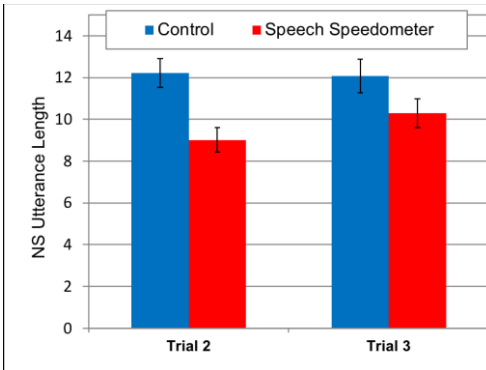


Figure 5: Mean NS Utterance Length (words per utterance) by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean).

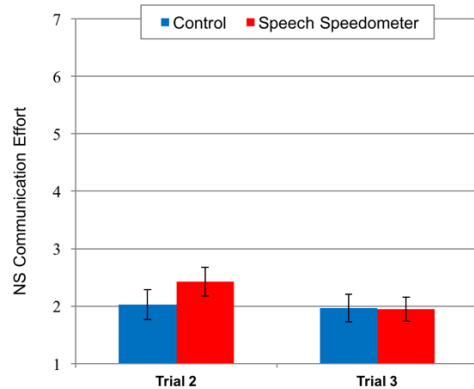


Figure 6: Mean NS Communication Effort on a scale of 1 to 7, by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean).

RQ2 asked if the Speech Speedometer interface would affect NS *workload* or *communication effort*. To explore this, we conducted 2x2 Mixed Models ANOVAs on NS *workload* and *communication effort*. No significant main effect of Speech Speedometer or trial was found for *workload*, suggesting that the Speech Speedometer interface did not pose greater *workload* on NS. There was however a significant main effect of trial ( $F [1, 38] = 6.98, p = .01$ ) on NS *communication effort*. NS reported exerting more *communication effort* (to follow, engage in and contribute to the conversation) in Trial 2 ( $M=2.23, SE=.26$ ), when the Speech Speedometer was displayed, than in Trial 3 ( $M=1.96, SE=.23$ ). This main effect was further qualified by a significant interaction effect between condition and trial ( $F [1, 38] = 4.00, p = .05$ ; see [Figure 6](#)), suggesting this significant difference only occurred for NS in the Speech Speedometer condition.

### 4.3 Effects of Speech Speedometer on NNS Perception of NS Speech and Other Subjective Ratings

4.3.1 *NNS perceived NS speech clarity*. H3a hypothesized that NS speech rate would be negatively associated with NNS perceptions of the clarity of NS speech. A Pearson's Correlation analysis showed no significant correlation between NS speech rate (regardless of condition) and NNS perceived NS speech clarity ( $r = -.16, p=.15$ ). There was also no significant correlation between NS utterance length (regardless of condition) and NNS perceived NS speech clarity ( $r=.03, p=.82$ ).

H3b hypothesized that NNS would perceive greater clarity of NS speech when their NS partners were given the Speech Speedometer. To test this, we conducted a 2x2 Mixed Models ANOVAs on NNS perceived clarity of NS speech. There was a significant main effect of Speech Speedometer ( $F [1, 38] = 4.35, p = .04$ ; see [Figure 7](#)) but in the opposite direction than predicted. No significant effect of trial ( $F [1, 38] = 1.75, p = .19$ ) and no interaction ( $F [1, 38] = 1.75, p = .19$ ) was found. Contrary to H3b, participants perceived lower clarity of NS speech when NS used the Speech Speedometer ( $M=5.28, SE=.33$ ) than when they used the control display ( $M=6.00, SE=.27$ ).

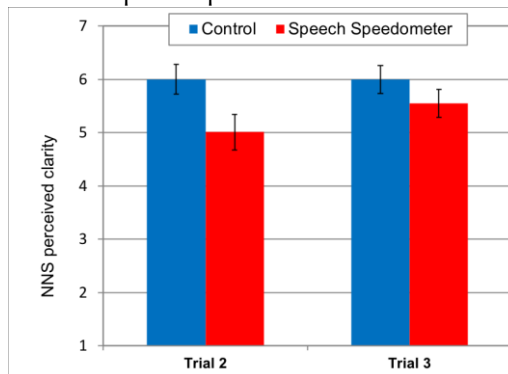


Figure 7: Mean NNS Perceived NS Speech Clarity on a scale of 1 to 7, by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean).

4.3.2 *NNS understanding of multiparty discussion.* H4a hypothesized that NS speech rate would be negatively associated with NNS understanding of multiparty discussion. A Pearson's Correlations analysis showed no significant correlation between NS speech rate (regardless of condition) and NNS understanding of multiparty discussion ( $r = .10$ ,  $p = .36$ ), and no correlation between NS utterance length (regardless of condition) and NNS understanding ( $r = .26$ ,  $p = .10$ ).

To test H4b that hypothesized NNS will have better understanding of multiparty discussion when NS were given the Speech Speedometer, we conducted Mixed Models ANOVAs on NNS understanding of multiparty discussion. There was no significant main effect of Speech Speedometer ( $F [1, 18] = 2.29$ ,  $p = .15$ ), no effect of trial ( $F [1, 18] = .91$ ,  $p = .35$ ) and no interaction ( $F [1, 18] = 1.68$ ,  $p = .21$ ). Contrary to H4b, NNS did not report significantly better understanding of multiparty discussion when NS were given the Speech Speedometer.

4.3.3 *NNS workload and communication effort.* H5a hypothesized that NS speech rate would be positively associated with NNS workload and communication effort. A Pearson's Correlations analysis showed no significant correlation between NS speech rate (regardless of condition) and NNS workload ( $r = -.08$ ,  $p = .49$ ), or between NS speech rate and communication effort ( $r = -.09$ ,  $p = .41$ ). There were also no significant correlations between NS utterance length (regardless of condition) and NNS workload ( $r = -.19$ ,  $p = .22$ ) or communication effort ( $r = .10$ ,  $p = .53$ ).

To test H5b, which hypothesized NNS would report lower *workload* and *communication effort* when communicating with NS who were given the Speech Speedometer, we conducted Mixed Models ANOVAs on NNS workload and communication effort. For workload, there was no significant main effect of Speech Speedometer ( $F [1, 18] = .40$ ,  $p = .54$ ), no effect of trial ( $F [1, 18] = .07$ ,  $p = .79$ ) and no interaction ( $F [1, 18] = 3.65$ ,  $p = .07$ ). Similarly, for communication effort we found no effect of condition ( $F [1, 18] = .86$ ,  $p = .37$ ), trial ( $F [1, 18] = .09$ ,  $p = .77$ ) or interaction ( $F [1, 18] = .09$ ,  $p = .77$ ). Contrary to H5b, NNS did not report significantly lower workload or communication effort when communicating with NS who were given the Speech Speedometer.

4.3.4 *NNS perceived NS accommodation.* H6a hypothesized that NS speech rate will be negatively associated with NNS perceived accommodation. A Pearson's Correlation analysis showed no significant correlation between NS speech rate (regardless of condition) and NNS perceived NS accommodation ( $r = -.07$ ,  $p = .52$ ). Nor was there a significant correlation between NS utterance length (regardless of condition) and NNS perception of NS accommodation ( $r = .12$ ,  $p = .30$ ).

To test H6b, which predicted that NNS would feel more accommodated by NS who were given the Speech Speedometer, we conducted a 2 by 2 Mixed Models ANOVA on NNS perceived

accommodation from NS. There was a significant main effect of Speech Speedometer ( $F [1, 38] = 6.31, p = .02$ ) but in the opposite direction than predicted. As shown in Figure 7, NNS perceived significantly less accommodation from NS who were given the Speech Speedometer ( $M=5.32, SE=.29$ ) than from NS given the control display ( $M=6.02, SE=.48$ ). There was also a significant main effect of trial ( $F [1, 38] = 4.22, p = .05$ ) suggesting that NNS perceived significantly less accommodation from NS in the second trial where the Speech Speedometer or its equivalent display was present ( $M=5.50, SE=.24$ ) than in the third trial where it was removed ( $M=5.86, SE=.22$ ). No interaction ( $F [1, 38] = .11, p = .74$ ) was found.

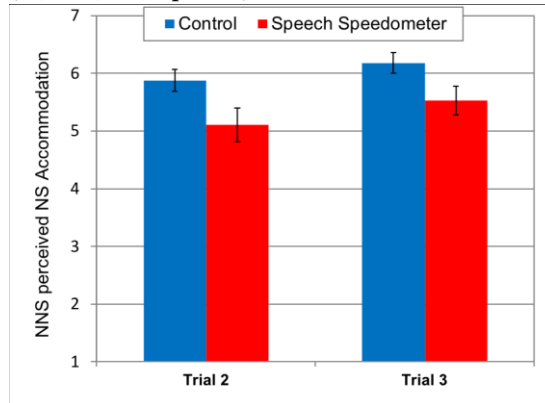


Figure 8: Mean NNS perceived NS Accommodation on a scale of 1 to 7, by presence or absence of Speech Speedometer for Trial 2 and Trial 3 (error bars represent standard errors of the mean).

#### 4.4 Insights from Interviews

All the interviews with NS were conducted in English, and were audio recorded and transcribed by the first author. All the interviews with NNS were conducted in Japanese, and were audio recorded and transcribed by the second author who is a native Japanese speaker. All the Japanese transcripts were translated into English by professional translators. Using an inductive and interpretive approach [24], we initially generated themes from the quotes, then identified high-level themes and relationships between the themes. We then went through an iterative process of generating and merging themes to refine the codes. Several interesting themes emerged from this analysis.

**4.4.1 Interviews with NS who were given the Speech Speedometer.** NS who were given the Speech Speedometer reflected on: whether and how exposure to the Speech Speedometer increased their awareness of the NNS partner's need for slowing down, whether and how NS used it to actually slow their speech, difficulties they experienced using the Speech Speedometer to slow their speech, and how they ended up modulating their speech in other ways.

The first theme that ran through interviews with all NS concerned their **awareness of the need to slow down**. Most NS who were given the Speech Speedometer clearly expressed that they knew slower speech would have been beneficial for NNS.

*You want to make sure that what you say is understood by both NS and NNS. And I think by adjusting one's speed, NNS can digest what you're saying more easily. – 2A*

*I think trying to keep the recommended low speed is better. I didn't really succeed to get down to smaller (slower). For her (NNS) it would have been helpful. I think*

*it's helpful for anybody when it's not their first language. Not as a reflection on their ability. Her English was quite good but I think it's helpful and respectful. – 3A*

*The thing that I learnt was, if I can speak at an appropriate speed maybe the vocabulary is not as important. I didn't want to use baby language. because I'm speaking about a very difficult topic and I can't dumb it down. I need my vocabulary to get some of this information out. – 1A*

Some NS indicated that they continued trying to slow down their speech even after the display was removed.

*The last (trial) I still tried to slow down. I was kind of imagining that graph. It was interesting how much I thought about the bars. – 1A*

*(In Trial 3), I was trying to slow down my speed or judge how fast I was speaking all the time. I wish there was a bar or something. – 9A*

The second theme concerned **whether and how NS used the Speech Speedometer to successfully slow their speech.**

Some found the Speech Speedometer effective in monitoring or adjusting their speech rate:

*The bar (Speedometer) controlled me. If you told me to speak slowly. I wouldn't know how slow to speak... or how slowly I am talking. For 20 years teaching English I thought just slowing down my speech is slow, but (no) this is slow. – 1A*

*It's like driving a car and you noticed you're speeding, well let's back off the gas paddle a little bit. It's helpful when you know the speed of your speech pattern. – 2B*

However, despite feeling more motivated to slow down in the Speech Speedometer condition, most NS reported difficulties of slowing their speech to the Speedometer.

Some difficulties derived from the infrequent update of the Speech Speedometer.

*It felt like if I slow down, the bar didn't respond immediately. It didn't feel like it was responding very quickly to my changes. If it could respond faster it would probably be more effective. – 4B*

*It would be better it would be able to show a faster change in speech rate. So that the person can see "I'm beginning to speak faster before I hit the red, lets slow down". – 6A*

Some NS admitted that it's difficult to slow down when NS dominate the conversation.

*We (2NS) tend to speed each other up... It's kind of easy for us to pingpong off of one another where we were speeding up, according to the bar. If we (2NS) were doing most of the talking to one another, it's very easy for us to fall into that rhythm, and maybe forget that there's a third person listening. – 4B*

Additionally, several NS mentioned that it was mentally taxing and distracting to focus on speed of talking:

*It felt really uncomfortable to try to maintain that speed, to keep within the green... I felt labored. – 6A*

*The effort in speaking that slowly would eventually become so distracting that... I would be slowing down to the point that start losing track of...maybe the same way you speak too quickly that you get lost in the speech, you can also get lost in slowness... It's like a mental exertion in slowing the ideas coming down. – 4B*

Others also noted that the speed suggested by the Speedometer was not appropriate:

*Even the average speed of 140 could seem a bit misleading because you can say something quickly, especially expressions that people know...I could be wrong but I felt like my self-regulation was doing a better job than the Speedometer. – 9B*

*I felt quite patronizing to consciously try to speak (that slow to) make it green. – 3B*

As a result, some gave up trying to lower their speed,

*(I tried to keep within the green) at first. But then I realized every time I spoke it would just gonna be in the red so I just stopped caring. – 6B*

While NS in the Speech Speedometer condition were generally motivated to slow down their speech, a few NS were not motivated to slow down. These NS seemed to ignore the Speech Speedometer and trust their own judgment of the NNS language ability rather than what the Speech Speedometer suggested.

*I just felt like she could understand, no matter how fast I was talking. So I didn't feel the need to adjust. I feel like I'm pretty good at judging their English level so if I need to slow down then I will – 6B*

*On the speed factor, with NNS that you know are struggling in listening...I try to speak more slowly when I know the person is not as proficient. But it felt like...more so in the second conversation and the third one, that C (NNS) was more aware of what we were talking about and able to at least gloss what we were saying. That awareness probably had us less concerned about slowing down. – 4B*

The last theme concerned how NS ended up **adding more pauses** as a way to slow down and how they justified their adjustment:

*You can say something quickly, especially expressions that people know, and then you pause to let what you just said sink in. Because it's not the speed of the... it takes time to think about what was said and (that would make them) miss ongoing conversation. – 5B*

*I made a lot of pauses, because I'm thinking just how is the simplest way to explain to him (NNS) what I am thinking rather than what's my first reaction. – 9B*

Longer pauses were also used by NS to allow NNS partners to enter the conversation.

*If I continue for 30 or 60 seconds with just a pause to inhale, when do you speak? You have an opportunity to interrupt if I pause more frequently. That's why he (NNS) could pull in. –1A*

*But just that B(NS) and I were working on a faster level, that split second (of pause) is a big difference because when she (NNS) wants to chime in...it was one second ago, and B and I are already talking about the next. We are half a sentence further.*



*She might have something very pertinent and insightful and knowledgeable, related to what we were talking about but realized her chance has passed. –4A*

4.4.2 *Interviews with NNS.* Interviews with NNS reveal several themes surrounding NNS perception of changes in NS speech rate, clarity, accommodation and attitude across trials, and how those changes influenced NNS own behavior and attitude in the communication.

Regarding the **perception of NS speaking speed**, 7 out of 10 NNS in the Speech Speedometer condition felt their NS partners spoke slower in Trial 2 and 3 than in Trial 1 (the warm up trial).

*The 1st time B was (spoke) very fast. From the 2nd time B listened to other's opinions and waited a lot. – 3C (Speedometer)*

*It might be due to the theme (topic), but I felt like B spoke a little more slowly and it was easier to understand in the 2nd and 3rd times. –9C*

*I feel like A spoke slower the 2<sup>nd</sup> time. The final time was the slowest. It was slow and easy to listen to. – 1C*

Whereas in the control condition, 5 out of 10 felt the conversation went faster and faster.

*The 1st time in particular they spoke very slowly. The 2nd and 3rd times they spoke a little more quickly. And the vocabulary became difficult. For the 1st time, we all didn't know each other, so we spoke slowly while observing the situation. For the 2nd time, they probably thought I understood, so they spoke a little faster. In particular for the last one, it took everything just to keep up and I couldn't speak. – 14C (control)*

*The more we got used to it, the intimacy increased, and I feel like the speed increased. –16C (control)*

With respect to the **perception of NS speech clarity and accommodation**, several of them reported feeling that NS “spoke in a way that was easier” for them to understand in Trial 2 and 3 than in Trial 1. NNS reported higher clarity and accommodation when NS used verbal accommodation techniques such as using simpler words or providing examples for NNS.

*After the first time, A gave examples to make things easier for me to understand. A used simple words and spoke slowly. They may have thought I didn't speak much during the first time. – 10C*

However, accommodation without explicit verbal indication (i.e. simply trying to slowdown to match the speed of the speedometer) was rather unnoticed by NNS. Or even if some of them did notice the change in their ease of understanding and following the conversation, they did not attribute it to NS deliberate effort to accommodate for them.

*(after the 1st trial) listening was easier. I could clearly take it what the other person was saying. Perhaps because it took them time to think, too. The 2nd (trial) in particular was easy to understand. –9C (Speedometer)*

*There were short breaks (pauses) which I took advantage of. I felt that they were talking slowly while thinking. I didn't feel like it was intentional. –5C (Speedometer)*

In fact, several NNS noted that they sensed “something was going on” (when the Speedometer was displayed) that made NS “hold back their speech”, which made their argument incomplete and confusing for NNS. It also made NNS feel their NS partner lost interest in the discussion.

*For the 2nd time I felt at one point that something wasn't quite right. They (NS) talked less and were holding back their opinions... In the discussion, the options C and D were quickly eliminated, but they didn't explain why. – 3C (Speedometer)*

*In the second discussion, I felt that something was going on for B. He wasn't paying attention. He lost passion. There was an increase of pointless conversation. – 7C (Speedometer)*

These subtle changes in NS speech behavior, particularly their pausing behavior, **increased the ease of speaking up** for NNS. Half of NNS in the Speedometer condition felt it was “easier to talk” in the 2<sup>nd</sup> trial. They felt NS waited more for them and held back their own speech to give them chance to speak.

*If they're just talking constantly it's difficult to jump in. I was happy they waited and listened and that made it easy to join. Midway, I felt that B really would have been talking, and that B held back in order to give me the chance to speak. –3C*

*The second time A let me speak a lot. A was being mindful, like A was holding back, listening intently to the other person's opinion and was trying to wait...the 2nd (trial) was the most enjoyable, that's where I could say the most of what I wanted to say. –4C*

However, this floor offering behavior of NS also made some NNS feel stressful because they felt they were put on spot and they didn't want to keep NS waiting while they formulate their thoughts in English.

*Having them ask my opinion made it easy to talk. However, I felt stress as they waited for me, I felt pressure. –3C*

*It was easy to jump in, but I felt pressure as I couldn't say anything important. –8C*

*The 1st time I didn't speak at all. It was easier to talk the 2nd time... (but) I didn't speak up soon, (I) rather tried to refine my thoughts internally. But then I couldn't express my refined thoughts in English as I would have liked. –9C*

## 5 DISCUSSION

Overall, our statistical analysis suggests that despite feeling more motivated to slow, NS provided with our Speech Speedometer did not successfully slow down their speech, although they did produce shorter utterances. However, it cost them more communication effort when they were given the Speech Speedometer. Contrary to our hypotheses, NNS perceived lower clarity of NS speech and reported lower perceived accommodation from NS when NS were given the Speech Speedometer. In the rest of this section, we draw on the literature and participants' reflections to interpret the findings.

### 5.1 Effects of Speech Speedometer on NS

The increase in NS motivation to slow when given the Speech Speedometer provides evidence that the design of the system was at least partially successful in achieving its goal of increasing NS

awareness of their speech rate. The interviews further suggest that most NS became mindful of their speech patterns and of the need to accommodate NNS. They continued to make effort to try to accommodate NNS partners even after the Speech Speedometer was removed.

There are several reasons why the motivation to slow down might not have translated into an actual drop in speech rate. Apart from the difficulties in getting feedback from the Speech Speedometer, which we will discuss further in design implications, NS suggested they had difficulty learning exactly *how* to moderate their speech to achieve the goal of slower speech rate. This is perhaps not surprising given that NS speech production is generally outside conscious attention (e.g., [45]). As several of the NS noted, a conscious focus on slowing down was mentally taxing and disruptive to other dimensions of their communication. Some NS appear to have focused on more frequent pausing, which could be easier than deliberate slowing of speech. These observations suggest that the Speech Speedometer may have only succeeded at enabling NS to monitor and attend to their speech rate pattern. Evaluative as it was, it still lacked the tutorial instructions as to what specific action is needed (e.g. lengthening vowels [51]) in order to achieve the recommended speech rate. This highlights the need for additional scaffolding to help NS understand how to realize their accommodation goals.

## 5.2 Effects of Speech Speedometer on NNS

In the interviews, NNS gave positive reflections on their communication experience in sessions where Speech Speedometer was given to NS and then removed. They found the conversation was easier to follow and that they could speak up more easily in between NS turns. This suggests that NNS did perceive the changes such as pausing, in NS speech behavior. However, some NNS did not interpret those changes as intentional or deliberately made for them.

On the quantitative side, NNS perceived lower clarity as well as lower accommodation when NS were given the Speech Speedometer. It's worth noting that despite the significant differences, perceived NS accommodation for the Speech Speedometer condition was about 5 on a 7 point scale, suggesting NNS still felt a fair amount of adjustment. That said, these findings are counterintuitive and inconsistent with literature on the effect of pauses on perceptual judgments of speech clarity [5]. Drawing on NS and NNS reflections in the interviews, we have a better understanding of what may have happened when NS were given the Speedometer. We suspect that the difficulties NS experienced moderating their speech might offer one possible explanation why NNS perceived lower clarity of NS speech. NS reported that trying to adjust to the Speedometer caused them to "lose track of the thought process". When the NS lost track of thought while proposing their argument, the incomplete line of argument may have created confusion for NNS. As one NNS noted, she couldn't understand why certain options were eliminated because the NS "was caught in something" and failed to give an explanation. Additionally, when making these sporadic attempts to respond to the Speedometer, NS sounded to NNS as if they were "holding back". Some NNS interpreted this behavior as NS losing interest in the discussion. This perception might explain why NNS perceived NS as less accommodating. Also, when attending to the Speedometer, NS might have seemed distracted to NNS. As some of them noted, they sensed that "something was going on" for the NS that "kept them from focusing on the task". These observations highlight the need for communicating NS effort of accommodation to NNS.

## 6 DESIGN IMPLICATIONS

In this section, we first discuss how the results from the current study can be directly used to design a better and smarter Speech Speedometer. Designing tools to help NS moderate their speech rate is only a small step towards a holistic approach to address communication issues in multilingual collaboration. We therefore draw on our findings to further suggest several ways that technology may facilitate multilingual collaboration between NS and NNS by implementing speech accommodation training, and enabling a meta-communication channel to align accommodation efforts and needs between NS and NNS.

### 6.1 A Better Real-Time Feedback System that Balances Frequency of Updates

To reach a balance between too frequent and too few updates, frequency rules can be applied on top of the determined silence threshold. For instance, if 0.5s threshold is used, the system could average a bunch of speech rates for adjacent utterances within a few seconds to generate one update for that duration. If 1.2s is used, the system could force to generate updates every few seconds for already transcribed utterances if the speaker continues to talk nonstop. In retrospect, we may also have set the parameter of “acceptable” speech rate too slow, which was too difficult for NS to achieve in conversation. Setting a wider range of acceptable speech rates might be easier for NS to achieve and make them feel more rewarded and motivated to try, as the likelihood to get positive feedback is higher. Additionally, with regard to the issue reported by some NS, that constantly focusing on speech rate could disrupt their thought process, a peripheral display of the Speedometer or even an alarm display that appears only when the NS speech rate exceeds the threshold can be used instead.

### 6.2 A Smarter Speech Speedometer Using NLP

Findings from both the current study and previous studies [11][57] have suggested that the cost and benefit of slower speech rate interacts with various factors, among them vocabulary difficulty (see quotes from Participant 9B for example) and frequency, syntax complexity etc. With advancements in natural language processing, we may be able to incorporate lexical frequency and syntax difficulty into a smarter Speech Speedometer that would automatically raise speech rate thresholds for difficult or less frequent words and phrases such as idioms and slang that a NNS would likely to be unfamiliar with, and lower the thresholds for common usage that wouldn't take NNS much mental resources to process. A smarter Speech Speedometer might also update speech rate targets regularly, even within a sentence, by detecting semantic completion rather than relying on acoustic silence. This way, the speedometer could help NS monitor and modify their speech rates in near real-time.

### 6.3 Scaffolding NS Accommodation

NS in our study had significant difficulty slowing their speech based on the Speech Speedometer, despite being aware of and motivated to work towards the goal. Besides improving the current system, we suggest that tools will need to provide additional scaffolding to help NS achieve their behavioral goals, such that their motivation can transform into actual behavior change [8][16]. For example, speech speed goals might be set by individual NS collaboratively with the system, so as to be challenging but realistically achievable. Alternatively, the goals might be introduced gradually, in small increments, so that the NS get immediate positive feedback. This would require less mental effort to achieve a phasic goal and NS might feel rewarded to move onto the next phase. This way, NS might not feel so discouraged by their poor performance at achieving an

ultimate speech rate goal as to give up trying immediately. Additionally, we might create tools along the lines of [62] that allow NS to practice speech accommodation outside conversational contexts.

## 6.4 Creating a Channel for Two-way Communication of NS Accommodation Effort and NNS Needs

Although NNS in our study generally perceived changes in NS speech behavior, some did not interpret those behavior changes as deliberately made for them. In other words, some NNS were not aware of that their NS partners made the effort to accommodate. On the other hand, some NS (e.g. Participant 6B) did not think accommodation was necessary for their NNS partner. They trusted their own judgement more than what the tool advised them to do, which was unwise as interviews with their individual NNS partners suggest they still needed NS to speak slower for them. These misinterpretations and misjudgments of both parties suggest a need for communication of NS accommodation effort to NNS and that of NNS specific needs to NS. Eye-tracking technology might be applied to measuring how much NS is attending to the Speedometer which is then presented to NNS using visual summary to indicate their accommodation effort. Advance speech analysis such as pause count, word difficulty analysis can be combined with eye-tracking to calculate a measure for NS accommodation effort for NNS to see. Alternatively, we suppose that enabling NNS to give feedback on NS speech rate regularly (e.g. approving the speed or showing appreciation) may not only allow NNS to see NS effort (or lack thereof), but also align the amount of accommodation with NNS needs. Matching speech adjustments to the needs of individual NNS and even the moment by moment needs of theirs would be ideal, but likely beyond the abilities of most NS. Also, previous research suggests that many NNS prefer not to be explicit about their needs for help, due to face concerns [10][13][46]. An alternative strategy might be to automatically slow speech and insert pauses when NS are talking, and combining this with other tools that can create the necessary gaps in the conversation for this process to occur (e.g., [69]). Another alternative could be to give NNS the options and control of the types of interventions at their disposal or through an agency. For instance, if NNS needed NS to slow down, they could do so with a click, and it either automatically slows down through silence gaps insertion, or signals NS to slow down by use of the Speedometer, or through the facilitation of a bot (e.g. [13]). In this way, NNS actual need for accommodation can be addressed without their face being threatened.

## 7 LIMITATION AND FUTURE WORK

The findings of the current study need to be interpreted with several limitations in mind. We discuss the limitations with respect to our design decisions, choice of task, and the representativeness of our sample. These limitations also lay out directions for future work to explore.

First, how often the Speech Speedometer gave feedback was largely dependent on NS pausing style. NS who tend to make more long pauses will get more frequent update on their speaking speed. The Speech Speedometer apparently was insensitive to the within-utterance adjustments as to be effective. As we have mentioned earlier in this paper, the frequency of speech rate update was a trade-off between frequency and meaningfulness. To avoid too frequent updates that caused NS confusion and distraction in pilot testing, we deliberately adjusted the silence threshold from 0.5s to 1.2s such that it generates feedback only when it detects 1.2s or longer no-talk. By doing this, we also intended for our speech rate measure to account for intra-utterance pauses. This

design decision ended up biasing those who naturally make more long pauses to receiving more frequent feedback on their speech rate.

Second, we deliberately chose the task to simulate real-world international collaboration, where NS tend to talk at their natural speed and be so immersed in the discussion as to forget to slow for NNS. While we achieved the goal of getting NS to speak fast using the task, it was also hard for them to slow down talking about those hot issues, even with the tool provided. Future research may adopt other types of task to investigate how characteristics of task affect the ease of slowing down. One example could be an instruction giving task where the goal of information transmission takes priority over expression of personal opinions.

Finally, being conducted in Japan, this study used an unrepresentative sample of NS participants, most of whom have lived in Japan for a long time and are used to talking to NNS. They may have had a greater tendency than most NS to accommodate for NNS. Being a NNS (of Japanese) themselves living in a foreign country, they may have more empathy for NNS of English than an average NS. That said, even this supposedly accommodating population failed to slow their speech to the extent that an intermediate level NNS would benefit significantly. This suggests the complexity and difficulty of designing tools that can trigger immediate changes in speaking behavior without additional scaffolds. This limitation also leaves room for future endeavors to explore how an average NS living in their native country would react to the evaluative speech rate feedback implemented for remote global videoconferencing.

## ACKNOWLEDGMENTS

This work was funded in part by NSF grant #1421929. We also thank the NTT development team for their technical support.

## REFERENCES

- [1] Simon Arnfield, Peter Roach, Jane Setter, Peter Greasley, and Dave Horton. 1995. Emotional Stress and Speech Tempo Variation. *Speech under Stress*, September: 13–15. Retrieved from [http://www.isca-speech.org/archive\\_open/sus\\_95/sus5\\_013.html](http://www.isca-speech.org/archive_open/sus_95/sus5_013.html)
- [2] Leslie M. Beebe and Howard Giles. 1984. Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language* 46: 5–32.
- [3] Sonja Biersack, Vera Kempe, and Lorna Knapton. 2005. Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. *Interspeech 2005*: 2401–2404. Retrieved from [http://www.iscaspeech.org/archive/archive\\_papers/interspeech\\_2005/i05\\_2401.pdf](http://www.iscaspeech.org/archive/archive_papers/interspeech_2005/i05_2401.pdf)
- [4] Eileen K. Blau. 1990. The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly* 24, 4: 746–753.
- [5] Ann R. Bradlow and Tessa Bent. 2002. The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America* 112, 1: 272–284. <https://doi.org/10.1121/1.1487837>
- [6] Ann R. Bradlow and Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 2: 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>
- [7] David B. Buller and R Kelly Aune. 1992. The Effects of Speech Rate Similarity on Compliance- Application of Communication Accommodation Theory. *Western Journal of Communication* 56: 37–53.
- [8] Sunny Consolvo, David W. McDonald, and James A. Landay. 2009. Theory-driven design strategies for technologies that support behavior change in everyday life. *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM.
- [9] Peter J. Costa. (2014). Truncated outlier filtering. *Journal of Biopharmaceutical Statistics*, 24(5), 1115–1129.
- [10] Pino Cutrone. 2014. A cross-cultural examination of the backchannel behavior of Japanese and Americans: Considerations for Japanese EFL learners. *Intercultural Pragmatics* 11, 1: 83–120. <https://doi.org/10.1515/ip-2014-0004>
- [11] Tracey M. Derwing. 1990. Speech rate is no simple matter: Rate adjustment and NS-NNS communicative success. *Studies in Second Language Acquisition* 12, 3: 303–313. <https://doi.org/10.1017/S0272263100009189>
- [12] Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. 2004. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04)*. ACM, New York, NY, USA, 614–623. DOI=<http://dx.doi.org/10.1145/1031607.1031713>

- [13] Wen Duan, Naomi Yamashita, Sun Young Hwang, and Susan R. Fussell. 2018. "Let Me Ask Them to Clarify If You Don't Want To" — A Clarification Agent for Nonnative Speakers. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*: 1–6.
- [14] Alan J. Feely and Anne-Wil Harzing. 2003. Language management in multinational companies. *Cross-Cultural Management: An International Journal* 10, 2: 37–52.
- [15] Stanley Feldstein and Ronald N. Bond. 1981. Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech* 24.4: 387–394.
- [16] B. J. Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December, pages. DOI: <https://doi.org/10.1145/764008.763957>
- [17] Susan Anne Fox and Howard Giles. 1996. Interability communication: Evaluating patronizing encounters. *Journal of Language and Social Psychology* 15, 3: 265–290.
- [18] Ge Gao, Naomi Yamashita, Ari Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of public vs. private automated transcripts on multiparty communication between native and non-native English speakers. *Proceedings of the 2014 ACM Conference on Human Factors in Computing Systems*: 843–852. <https://doi.org/10.1145/2556288.2557303>
- [19] Ge Gao, Naomi Yamashita, Ari Hautasaari, and Susan R. Fussell. 2015. Improving multilingual collaboration by displaying how non-native speakers use automated transcripts and bilingual dictionaries. In *Proc. CHI 2015*, 3463–3472. <https://doi.org/10.1145/2702123.2702498>
- [20] Jessica Gasiorek. 2015. Perspective-Taking, Inferred Motive, and Perceived Accommodation in Nonaccommodative Conversations. *Journal of Language and Social Psychology* 34, 5: 577–586.
- [21] Howard Giles and Philip M. Smith. 1979. Accommodation theory: Optimal levels of convergence. *Language and social psychology*: 45–65.
- [22] Howard Giles and Tania Ogay. 2007. Communication Accommodation Theory. In *Explaining Communication: Contemporary Theories and Exemplars*, Bryan B. Whaley and Wendy Samter (eds.). 293–310. <https://doi.org/10.4324/9781410614308>
- [23] Howard Giles, Donald M. Taylor, and Richard Bourhis. 1973. Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data. *Language in Society* 2, 2: 177–192.
- [24] Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: strategies for qualitative research*. Aldine Publishing Company, Hawthorne, NY.
- [25] Frieda Goldman-Eisler. 1956. The determinations of the rate of speech and their mutual relations. *Journal of Psychological Research* 2: 137–143.
- [26] R. Griffiths. 1990. Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning* 40, 3: 311–336.
- [27] Francois Grosjean and Harlan Lane. 1976. How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance* 2.4: 538.
- [28] Sandra G. Hart and Lowell E. Staveland. 1998. Development of NASA-TLX: Results of empirical and theoretical research. *Advances in Psychology* 52: 139–183.
- [29] Jake Harwood and Howard Giles. 1996. Reactions to older people being patronized: The roles of response strategies and attributed thoughts. *Journal of Language and Social Psychology* 15, 4: 395–421.
- [30] Anne-Wil Harzing and Alan J. Feely. 2008. The language barrier and its implications for HQ-subsidary relationships. *Cross Cultural Management: An International Journal* 15, 1: 49–61.
- [31] Ari Hautasaari and Naomi Yamashita. 2014. Do automated transcripts help non-native speakers catch up on missed conversation in audio conferences?. In *Proceedings of the 5th ACM international conference on Collaboration across boundaries: culture, distance & technology (CABS '14)*. ACM, New York, NY, USA, 65–72. DOI: <https://doi.org/10.1145/2631488.2631495>
- [32] Helen Ai He, Naomi Yamashita, Chat Wacharamanotham, Andrea B. Horn, Jenny Schmid, and Elaine M. Huang. 2017. Two Sides to Every Story: Mitigating Intercultural Conflict through Automated Feedback and Shared Self-Reflections in Global Virtual Teams. *PACM Human-Computer Interaction CSCW, Article 51 (December 2017)* 1: 21 pages. <https://doi.org/DOI:https://doi.org/10.1145/3134686>
- [33] Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4: 555–568. <https://doi.org/10.1016/j.wocn.2010.08.002>
- [34] Pamela J. Hinds, Tsedal B. Neeley, and Catherine Durnell Cramton. 2014. Language as a lightning rod: Power contests, emotion regulation, and subgroup dynamics in global teams. *Journal of International Business Studies* 45, 5: 536–561. <https://doi.org/10.1057/jibs.2013.62>
- [35] Yoshito Hirozane. 2012. Perception of English Rate by Japanese L2 Learners of English. Doctoral dissertation. Macquarie University.
- [36] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: my automated conversation coach. In *Proceedings of the 2013 ACM international joint conference*

- on Pervasive and ubiquitous computing (UbiComp '13). ACM, New York, NY, USA, 697-706. DOI: <https://doi.org/10.1145/2493432.2493502>
- [37] George Jacobs. 1988. The Effect of Pausing on Listening Comprehension.
  - [38] Jane Kassis Henderson. 2005. Language diversity in international management teams. *International Studies of Management and Organisation* 35, 1: 66–82. <https://doi.org/10.1080/00208825.2005.11043722>
  - [39] Ken Kelch. 1985. Modified input as an aid to comprehension. *Studies in Second Language Acquisition* 7, 1: 81–90.
  - [40] A. Klueger,, A.DeNisi, 1996. Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119.2: 254-284.
  - [41] Jacques Koreman. 2006. Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *The Journal of the Acoustical Society of America* 119.1: 582-596.
  - [42] Jean C. Krause and Louis D. Braid. 1995. The effects of speaking rate on the intelligibility of speech for various speaking modes. *The Journal of the Acoustical Society of America* 98, 5: 2982–2982.
  - [43] Norman J Lass. 1970. The significance of intra-and intersentence pause times in perceptual judgments of oral reading rate. *Journal of speech and Hearing Research* 13.4: 777-784.
  - [44] Gilly Leshed, Diego Perez, Jeffrey T. Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L. McLeod, and Geri Gay. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 537-546. DOI: <https://doi.org/10.1145/1518701.1518784>
  - [45] Willem J. Levelt 1993. Speaking: From intention to articulation (Vol. 1). MIT press.
  - [46] Han Z. Li, Young-ok Yum, Robin Yates, Laura Aguilera, Ying Mao, and Yue Zheng. 2005. Interruption and Involvement in Discourse: Can Intercultural Interlocutors be Trained? *Journal of Intercultural Communication Research* 34, 4: 233–254.
  - [47] Na Li and Mary Beth Rosson. 2012. At a different tempo: What goes wrong in online cross-cultural group chat? *2012 17th ACM International Conference on Supporting Group Work, GROUP 2012*: 145–154. <https://doi.org/10.1145/2389176.2389200>
  - [48] Joseph Edward McGrath. 1984. *Groups: interaction and performance*. Prentice-Hall, Englewood Cliffs, NJ.
  - [49] Albert Mehrabian and Martin Williams. 1969. Nonverbal concomitants of perceived and intended persuasiveness. *Journal of Personality and Social Psychology* 13, 1: 37–58.
  - [50] Brian E. Mennecke, Joseph S. Valacich, and Bradley C. Wheeler. 2000. The Effects of Media and Task on User Performance: A Test of the Task-Media Fit Hypothesis. *Group Decision and Negotiation* 9, 6: 507–529. <https://doi.org/10.1023/A:1008770106779>
  - [51] Joanne L Miller, Grosjean, François, and Concetta Lomanto. 1984. Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*. 41.4: 215–225.
  - [52] Mei-hua Pan, Naomi Yamashita, and Hao-chuan Wang. 2017. Task Rebalancing: Improving Multilingual Communication with Native Speakers - Generated Highlights on Automated Transcripts. *Proc. CSCW 2017*: 310–321. <https://doi.org/10.1145/2998181.2998304>
  - [53] Yingxin Pan, Danning Jiang, Michael Picheny, and Yong Qin. 2009. Effects of real-time transcription on non-native speaker's comprehension in computer-mediated communications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2353-2356. DOI: <https://doi.org/10.1145/1518701.1519061>
  - [54] Hafiyah Pratiyanto, Takashi Nose, Yuya Chiba, and Akinori Ito. 2018. Analysis of preferred speaking rate and pause in spoken easy Japanese for non-native listeners. *Acoustical Science and Technology* 39.2: 92-100.
  - [55] C. M., Rietveld, & Carlos Gussenhoven. 1987. Perceived speech rate and intonation. *Journal of Phonetics*, 15.3: 273-285.
  - [56] Pamela Rogerson-Revell. 2008. Participation and performance in international business meetings. *English for Specific Purposes* 27, 3: 338–360. <https://doi.org/10.1016/j.esp.2008.02.003>
  - [57] Joan Rubin. 1994. A Review of Second Language Listening Comprehension Research. *The Modern Language Journal* 78, 2: 199–221.
  - [58] Lu-Feng Shi and Nadia Farooq. 2012. Bilingual Listeners' Perception of Temporally Manipulated English Passages. *Journal of Speech, Language, and Hearing Research*. 55: 125–138
  - [59] Richard L Street Jr., 1982. Evaluation of noncontent speech accommodation. *Language & Communication* 2, 1: 13-31.
  - [60] Emma Sweeney and Zhu Hua. 2010. Accommodating toward your audience: Do native speakers of English know how to accommodate their communication strategies toward nonnative speakers of english? *Journal of Business Communication* 47, 4: 477–504. <https://doi.org/10.1177/0021943610377308>
  - [61] Yohtaro Takano and A. Noda. 1993. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology* 24, 4: 445–462. <https://doi.org/10.1177/0022022193244005>



- [62] M Iftekhhar Tanveer, Emy Lin, and Mohammed (Ehsan) Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *IUI 2015*.
- [63] Steve Tauroza and Desmond Allison. 1990. Speech Rates in British English. *Applied Linguistics* 11, 1: 90–105.
- [64] Jurgen Trouvain. 2003. Tempo Variation in Speech Production Implications for speech synthesis.
- [65] Hao-chuan Wang and Susan R Fussell. 2010. Groups in Groups: Conversational Similarity in Online Multicultural Multiparty Brainstorming. *ACM conference on Computer supported cooperative work*: 351–360. <https://doi.org/10.1145/1718918.1718980>
- [66] Ann Wennerstrom and Andrew F. Siegel. 2003. Keeping the floor in multiparty conversations: intonation, syntax, and pause. *Discourse Processes* 36, 77–107. [https://doi.org/10.1207/S15326950DP3602\\_1](https://doi.org/10.1207/S15326950DP3602_1)
- [67] Blue Wooldridge. 2001. “Foreigner talk”: An important element in cross-cultural management education and training. *International Review of Administrative Sciences* 67, 4: 621–634. <https://doi.org/10.1177/0020852301674002>
- [68] Jean. Wong. 2000. The token “yeah” in nonnative speaker english conversation. *Research on Language and Social Interaction* 33, 1: 39–67.
- [69] Naomi Yamashita, Andy Echenique, Toru Ishida, and Ari Hautasaari. 2013. Lost in Transmittance: How Transmission Lag Enhances and Deteriorates Multilingual Collaboration. *ACM Conference on Computer Supported Cooperative Work Social Computing Social Computing*: 923–934. <https://doi.org/10.1145/2441776.2441881>
- [70] Jing Ye, Tomoo Inoue, A speech speed awareness system for non-native speakers, *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW'16 Companion)*, pp.49-52, 2016.
- [71] Chien Wen Yuan, Leslie D. Setlock, Dan Cosley, and Susan R. Fussell. 2013. *Understanding Informal Communication in Multilingual Contexts*. In *CSCW'13*.

Received April 2019; revised June 2019; accepted August 2019