What is the appropriate speech rate for a communication robot?

Michihiro Shimada & Takayuki Kanda
ATR, Japan

This study investigates the influence of a robot's speech rate. In human communication, slow speech is considered boring, speech at normal speed is perceived as credible, and fast speech is perceived as competent. To seek the appropriate speech rate for robots, we test whether these tendencies are replicated in human-robot interaction by conducting an experiment with four rates of speech: fast, normal, moderately slow, and slow. Our experimental results reveal a rather surprising trend. Participants prefer normal and moderately slow speech to fast speech. A robot that provides normal or moderately slow speech is perceived as competent. We further study how context affects this perception. In a situation where the robot and participants talk while walking, we found that slow speech was the most comprehensible. In addition, slow speech is subjectively perceived as good as moderately slow and normal speech.

Keywords: Human-robot interaction; speech rate

1. Introduction

Robots are being considered for information-providing roles. For instance, a museum-guide robot (Thrun et al. 1999) interacted with visitors in a human-like way using facial expressions and gestures. In an exposition (Siegwart et al. 2002) as well as in a supermarket (Gross et al. 2009), robots offered guided tours providing information while walking with visitors. In a shopping mall, a robot waits for visitors (Kanda et al. 2010) or proactively approaches them (Satake et al. 2009) to provide directions and information about shops. Since one strength of a 'robot' is its mobility, one typical scenario is where people are walking or standing after one of them has approached the other.

To make robot communication with people effective and efficient, researchers have studied how to use a robot's physical properties. Typically, the role of gaze and gesture in human communication is replicated in human-robot interaction. Robot

Interaction Studies 13:3 (2012), **408–435**. DOI 10.1075/is.13.3.05kan

gazes and gestures are useful in conveying intention (Breazeal, Kidd, Thomaz, Hoffman & Berlin 2005; Sidner, Lee, Kidd, Lesh & Rich 2005). Gaze was found to be an effective way to control listeners' attention (Mutlu, Forlizzi & Hodgins 2006). The standing position (Yamaoka, Kanda, Ishiguro & Hagita 2010) and spatial formation (Kuzuoka, Suzuki, Yamashita & Yamazaki) have been modeled from humans as well.

Moreover, recent studies have started to explore the temporal aspect for effective and efficient communication. For instance, pause attracts attention (Kuzuoka et al. 2008). Natural interaction requires an appropriate pause length (Yamamoto & Watanabe 2006) and response time (Shiwa, Kanda, Imai, Ishiguro & Hagita 2008). Okuno et al. reported that information presented at an appropriate speed facilitated listener understanding of route directions; however, that study was specific to a scenario of providing route direction (Okuno, Kanda, Imai, Ishiguro & Hagita 2009).

This study explores one factor related to such a temporal aspect: speech rate. Our first purpose is to reveal the influence of the robot's speech rate on comprehension as well as on the perceived impression. We also studied a related factor. Since robots have been typically used in either walking or standing situations, we explored whether a robot needs to adjust its speech rate to reflect this context.

2. Related works

A. Theory in human communication

Speech rate is an important factor in human communication. People speak at 125–150 words per minute (WPM) in normal conversations (Fulford & Zhang 1993). Faster speech (181 WPM) provides an extroverted impression, and speech with standard speed (154 WPM) provides a better impression of composure (Woodall & Burgoon 1983). Slower speech is boring (Johnstone & Scherer 2000), and faster speech is perceived as competent and dominant (Buller, Poire, Aune & Eloy 1992). Standard and faster speech are perceived as intelligent and credible (Simonds, Meyer, Quinlan & Hunt 2006). One study shows that in advertisement tasks, the standard speech rate (113 WPM) outperforms slower (56 WPM) or faster (212 WPM) speech rates (Goodman, Robinson, Robinson, Skinner & Sterling 1999). In Japanese, fast speech is considered less persuasive, so one should speak at normal speed to be persuasive (Tasaki 2003). Even though disagreement remains over whether faster speech is better than normal speech, the literature agrees that slower speech is boring and less persuasive.

Studies have also investigated the effect of speech rate on comprehension. People read at approximately 250–300 WPM, which is faster than their average speaking speed (Fulford 2001). Discussions have addressed whether people can comprehend speech at fast rates that exceed 250–300 WPM. Paul et al. found that although people were able to comprehend reasonably well until 300 WPM (but not at 375 WPM), their comprehension was better at 225 WPM (1992). Tun et al. found that people recall sentences less when speech is fast (at 182 and 280 WPM) or during a concurrent task that divided their attention (1992). Moreover, they reported that the decrease in sentence recall was lowest in normal speech (140 WPM). In contrast, at a medium rate of speed between 116–213 WPM, Simonds et al. found no significant effect of speech rate on recall (2006).

Note that in the above studies with human speakers, a speaker was typically asked to speak at normal, slow, or fast speeds, and the papers measured the WPMs of their speech. Thus, a speech rate categorized as "fast" in one study was labeled "normal" in another study (e.g. Woodall & Burgoon 1983 consider 181 wpm as fast, whereas it is considered a medium rate of speed in Simonds et al. 2006). We believe that this is partly due to the fact that the definition of fast speech depends on the speech contents.

B. Related research in spoken dialogs

The temporal aspect of speech has been extensively examined in speech synthesis studies. In speech synthesis, speech rate, which is included in computation as well as vowel and syllable pauses, is modeled from human speech (Zellner 1994).

Comprehension is the major target of such studies (Pisoni, Nusbaum & Greene 1985). When people first hear a synthesized voice they experience some comprehension difficulty, but comprehension improves after they get accustomed to listening to the synthesized voice, especially after the first five sentences (Venkatagiri 1994). In a study of the effect of slow speed in a synthesized voice, contrary to the authors' initial expectations, fast speech (210 WPM: originally labeled "normal" but we changed the designation for consistency with other studies) outperformed normal speech (150 WPM) by approximately 5% for word identification (Roring, Hines & Charness 2007). Their study also showed that natural speech was comprehended better than synthesized speech, and a younger population comprehended better than an older population.

In human communication, speech rate similarity between speakers and listeners provides better social attractiveness (Buller et al. 1992). To reproduce an interaction with speech rate similarity, studies have focused on mechanisms to measure user speech rates (Takamaru, Hiroshige, Araki & Tochinai 2000) and to pace synthesized voices to match user speech rates (Iwase & Ward 1998; Ward & Nakagawa 2002). Another study argues that task structure rather than the partner's speech rate is the dominant factor in human conversation and determines the speaking rate (Ward & Mamidipally 2008).

A few studies have investigated the effects of voice type and situation on speech comprehension. Females tend to provide better intelligibility, although individual differences are a much larger factor (Bradlow, Torretta & Pisoni 1996). Under highly noisy conditions, female voices provide less intelligibility due to acoustic differences (Nixon et al. 1998), although gender difference is considered negligible here (Edworthy, Hellier & Rivers 2003). When listening to a non-native language, a higher speech rate resulted in lower speech intelligibility but did not affect the results for persuasion (Jones, Berry & Stevens 2007). Listening in a noisy situation resulted in less comprehension, and elderly people were more affected; people tend to use more visual cues for comprehension in such noisy environments (Larsby, Hällgren, Lyxell & Arlinger 2005).

C. Previous studies in HRI

A number of research works have been carried out to build theoretical knowledge on natural interaction for social robots. How to use the physical properties of robots in interaction has been frequently studied, including the use of gaze (Mutlu et al. 2006; Rich et al. 2010; Sidner et al. 2005) and pointing (Brooks & Breazeal 2006; Kuzuoka, Oyama, Yamazaki, Suzuki & Mitsuishi 2000).

In contrast, the temporal aspect of social interaction remains relatively unexplored. Some studies have looked into turn-taking structure (Raux & Eskenazi 2009), pauses (Endrass, Rehm, André & Nakano 2008; Kuzuoka et al. 2008), and the timing of turn-taking (Robins, Dautenhahn, Boekhorst & Nehaniv 2008; Shiwa, Kanda, Imai, Ishiguro & Hagita 2009), but studies on speech rate are scant.

More specific to speech rate, as early as 1990, Watanabe studied a mechanism to adjust a robot's speaking rate to fit a user's speaking rate (1990). Scheutz et al. used speech rate to control the impressions of a robot (Scheutz, Schermerhorn & Kramer 2006). We also studied the timing of utterances (Okuno et al. 2009), but this work was specific to a given route direction. Aside from these exceptions, only a few studies have addressed the speech rate of robots, and none has addressed the effects of context on robots (e.g. the fact that a robot walks with people).

3. Experiment 1

In this paper, two experimental situations are reported: standing and walking. In the first experiment, we investigate which speech rate is appropriate when a robot and a participant talk while standing.

A. Hypothesis and predictions

Since the phenomena related to the listening process with robots are highly unexplored, we established three contradictive hypotheses based on different theories and/or considerations.

Hypothesis 1a: Robots should not speak more slowly than normal

This hypothesis is mainly derived from human communication literature (see Section 2.A for details). These studies agree about the usefulness in normal and fast speech and do not cite the need for slower speech. Thus, we made the following predictions:

- A robot that provides fast speech will be perceived as more competent than a robot that provides normal and slow speech (as Buller et al. 1992 reported for human communication).
- A robot that provides normal speech will be perceived as more credible than a robot that provides normal and slow speech (as Simonds et al. 2006 reported for human communication).
- Normal and fast speech will provide better comprehension than slow speech (as Pisoni et al. 1985 and Simonds et al. 2006 reported for human communication and Roring et al. 2007 reported for synthesized voices).
- Overall, normal and fast speech will be perceived as a better speech rate than a slow speech rate.

Hypothesis 1b: Robots should speak moderately slowly, but not too slowly

Contrary to the literature in human communication, we have empirically heard the opposite preference. When we conducted field trials (for example, Kanda et al. 2010), typical complaints concerned the speech rate of the robots. People claimed difficulty in understanding what the robots were saying, even after considerable habituation (Venkatagiri 1994).

During social interaction with robots, perhaps people experience a considerable extra cognitive load. They might need to observe the robot's motion, its behavior, its intention, and then consider appropriate social responses, all of which might increase their cognitive load. A study on extra cognitive load due to dual tasks (Tun et al. 1992) reported that a normal speech rate was better for comprehension than a fast one. They did not test a speech rate that is slower than normal speed. But we infer that under an extra cognitive load, slower outperforms normal.

Concerning the slowness of speech, the question is whether we should restrict slowness. Hypothesis 1b chooses to restrict the speech rate. Since the literature indicates no positive effect from slow speech, we should not make it too slow. Thus, in hypothesis 1b, we hypothesize that people prefer moderately slow speech (but not slow speech). The following predictions were made:

– A robot that speaks with a moderately slow speed will be perceived as more competent and credible than a robot using fast, normal, or slow speech rates.
– The robot's moderately slow speech will yield better comprehension than fast, normal, and slow speech rates.
– Overall, moderately slow speech will be perceived as a better speech rate than fast, normal, or slow speech rates.

Hypothesis 1c: Robots should speak as slowly as possible

This alternative hypothesis reflects a similar viewpoint to hypothesis 1b, which argues for slower speech. Hypothesis 1c argues that we do not need to restrict the slowness; we can make its speech rate as slow as possible within the range of the speed of natural human speech. It argues that people prefer slow speech.

Based on this consideration, the following predictions are made:

– A robot's slow speech will permit better comprehension than fast, normal, or moderately slow speech rates.
– Robots that speak at a slow speed will be perceived as more competent and credible than a robot using fast, normal, or moderately slow speech rates.
– Overall, slow speech rate will be perceived as better than fast, normal, or moderately slow speech rates.

B. Conditions

This experiment uses one factor, speech rate, and four conditions: fast, normal, moderately slow, and slow. The speech was provided in Japanese. Since adapting an English-based method for measuring words per minute to a system for use with Japanese is difficult, based on the literature, we investigated the standard speech rate in Japanese. Researchers used moras, a kind of syllable in Japanese, for measurement. Ward et al. reported that the Japanese speaking rate ranged from 6–10 moras/sec (Ward & Mamidipally 2008), where 8 moras/sec is considered standard (Yoshihara, Tsutaki & Takahashi 2008). Based on this, we set the following conditions:

- Fast (9.70 moras/sec): robot speaks around the range of the fastest speed when people speak fast.
- Normal (7.87 moras/sec): robot speaks at a standard speed of human speech.
- Moderately slow (6.93 moras/sec): robot speaks at a slower speed in the middle of the range where people speak more slowly.
- Slow (5.72 moras/sec): robot speaks at the slowest speed when people speak slowly.

Due to the parameter resolution that the speech synthesis software can accept, the moras per second are slightly different from the human standard. The speed of fast is approximately 20% faster than normal. This 20% difference resembles studies in the literature (Woodall & Burgoon 1983), but it is smaller than the parameters in other works (in Roring et al. 2007, fast was 40% faster than normal; in Goodman et al. 1999, twice/half speed to normal speed was used). Moderately slow was added between normal and slow because in a preliminary experiment participants preferred such an intermediate speed.

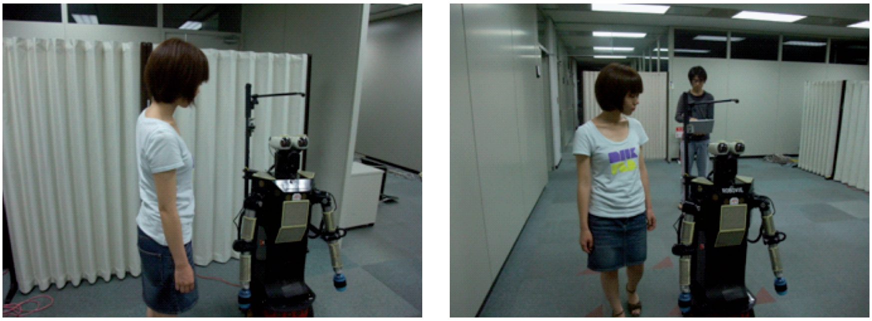
C. Method

1. *Participants*

Twenty-eight native Japanese speakers (17 males, 11 females, average age: 26.8) were paid for their involvement.

2. *Robot system*

We used Robovie-II, a 1.2-m tall communication robot with a 0.5-m radius, a 3-DOF head, and 4-DOF arms (Kanda, Ishiguro, Imai & Ono 2004). It moves on wheels and was controlled by the experimenter and provided Japanese utterances from a speaker attached to its head. Speech synthesis was performed using XIMERA software (Kawai, Toda, Ni, Tsuzaki & Tokuda 2004). We used a female voice. The synthesis software is a corpus-based speech synthesis that uses a large amount of recorded speech; when a fast/slow speech rate is specified, it searches for a segment that was spoken fast in the database to generate fast/slow speech as humans do.



(a) While standing (b) While walking

Figure 1. Robot talking with a person

3. *Setting*

The experiment was conducted in a laboratory room. When the robot provided speech, it was standing still and maintained eye contact with the participants, who listened to it (Figure 1(a)).

4. *Procedure*

People have difficulty hearing synthesized voices, particularly at first (Venkatagiri 1994). Accordingly, our participants became habituated to the robot's synthesized voice by listening to it tell a story for about four minutes.

Before starting the main sessions, we adjusted the volume (loudness) of its voice. Participants listened to a short sentence and adjusted the volume until they reported that it was appropriate. There was one practice session before the main session.

In the main session, participants engaged in story-telling type interaction with the robot, assuming that they are visitors to an information kiosk and seeking more information about sightseeing locations related to a place (e.g. Wyoming, Cuba, Sedona, Bolivia, Borneo, Mallorca, Mexico and Hangzhou). Before the session, participants were instructed to ask about a particular city or country, assuming that they are planning to visit it on their next vacation. When they met the robot, they asked it for information about a specific place that was provided by the experimenter. Then the robot started to give information about the location by explaining the name, some brief information about the available forms of transportation, the sightseeing spots in the area, including historical sites and wildlife attractions. It spoke about 20 sentences that lasted about 160 seconds at a normal speech rate.

We asked participants to relax and behave as visitors at a tourist information center. Although our post-session measurement asked them to provide the information they remembered, we explicitly informed them that they did not need to remember information for the test; otherwise, people would have tried to concentrate exclusively on remembering every word from the robot, which would have greatly altered the nature of the interaction which we wanted to observe.

The experiment was conducted as a within-subject design. Participants repeated all levels of speech rates twice to balance the possible deviation from individual differences in knowledge about particular locations. Eight experimental sessions were conducted. The measurements for the same speech rates were averaged for analysis, and the order of the sessions was counter-balanced. After participants had finished the listening sessions, they answered a questionnaire (details given in Section 3.D).

5. *Stimulus*

The robot's task was to provide information to users trying to find sightseeing locations that were selected based on their likelihood of being unfamiliar to ordinary Japanese people. This avoided any influence from previous knowledge participants might possess about these locations. If they knew a particular location well, listening to such information would be much easier and perhaps boring.

We controlled the difficulty and duration of the contents within the eight sightseeing locations offered by the robot. All contents consisted of approximately 20 sentences, each of which had approximately 43 moras. The difficulty was adjusted using preliminary trials based on comprehension test scores (explained in Section 3.D).

We made the speech as natural as possible. The sentences were prepared in a spoken language that imitated people's way of talking about information. Each sentence was followed by a short pause of 700 ms, which is a common duration when people speak (Jaffe & Feldstein 1970; Nagaoka et al. 2003).

D. Measurements

Three types of measurements were made: comprehension, subjective evaluation of speech rate, and subjective impression.

We measured participant comprehension:

– Comprehension: 10-point scale

Participants wrote what information they comprehended in their sightseeing plans. Since we asked them to include as many facts as possible, this test also has the characteristic of sentence recall. However, since without recall they cannot use it for planning their travel, we do not consider it possible to separate the recall factor as a measure of comprehension in sightseeing information. As spoken language comprehension is a complex interaction process in, measuring comprehension is difficult (Pisoni et al. 1985). Our test fitted the context of the information provided, where the participants' capability of providing a sightseeing plan indicated the degree of comprehension.

Evaluators who were blind to the experimental condition rated them on a 10-point scale based on the amount of facts from the robot's speech that were included in their written plans. Part of the data was evaluated by two evaluators; the inter-evaluator reliability (Cohen's weighted Kappa with quadratic weights) was 0.812, which indicates reasonable reliability.

We used a questionnaire on a 1-to-7 (low to high) scale to measure the participant's subjective ratings of the speech rate:

– Subjective evaluation of speech rate: 1–7.

The questionnaire asked participants to rate the overall goodness of the speech rate on a scale from 1-to-7 (bad and good).

We measured subjective impressions of participants' to the robot. Such impressions are known to be affected by speech rate (Buller et al. 1992; Simonds et al. 2006):

– Competence: 1–7 (incompetent to competent)

The questionnaire asked participants to rate the robot as a sightseeing guide on a scale from 1–7, where 1 is incompetent and 7 is competent.

– Credibility: 1–7 (not credible to credible)

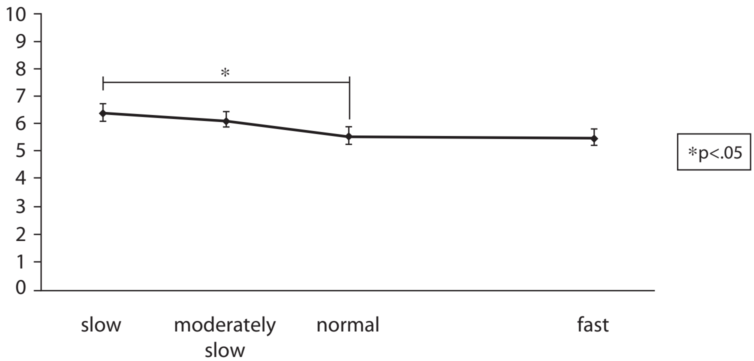
The questionnaire asked participants to rate the robot's talk on a scale from 1–7, where 1 is not credible and 7 is credible.

E. Results

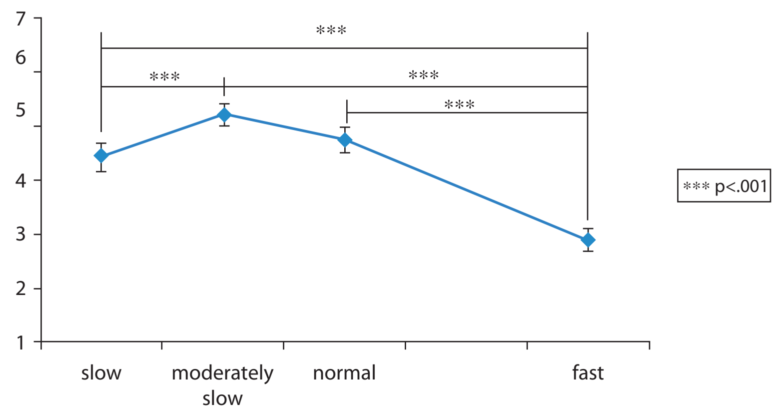
Figure 2 shows the result for each measurement. For multiple comparisons we used the Ryan method (1960), which is a step-down procedure that controls the family-wise error rate so that it is less than the alpha level (i.e. overall type-I error after repeated comparisons is less than 5%).

Comprehension: Figure 2(a) illustrates the comprehension results. The scores were 6.50, 6.08, 5.54, and 5.60 at slow, moderately slow, normal and fast speeds (SDs were 2.01, 1.52, 1.66, and 1.63). A one-way repeated measures analysis of variance (ANOVA) was conducted, and a significant main effect was found ($F(3,81) = 4.093$, $p = .009$, $\eta_p^2 = .132$). Multiple comparisons revealed significant differences between slow and normal ($p = .0110$). Comparisons between slow and moderately slow ($p = .333$) and moderately slow and normal ($p = .583$) were not significantly different. Other comparisons were far from significant ($p > .200$), except for the comparison between slow and fast, which approaches significance ($p = .0987$).

Therefore, for comprehension, the prediction in hypothesis 1c is supported. The prediction in hypothesis 1b is also supported, but with a caution, since slow was found to be better than normal but it was not concluded whether moderately slow is better than normal or worse than slow.

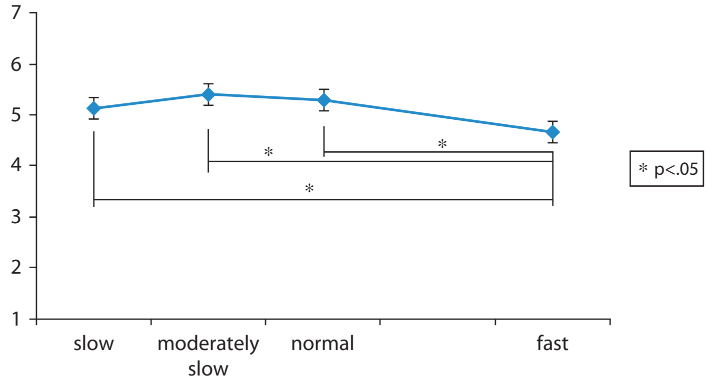


(a) Comprehension

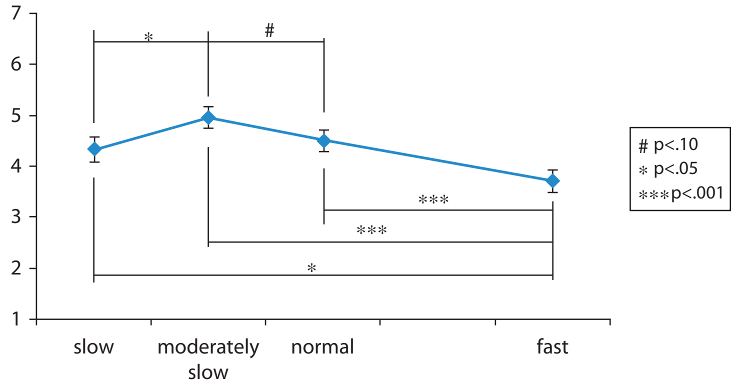


(b) Subjective evaluation of speech rate

Figure 2. Results from Experiment 1: standing situation *(Continued)*



(c) Credibility



(d) Competence

Figure 2. Results from Experiment 1: standing situation

Subjective evaluation of speech rate (Figure 2(b)): The scores were 4.45, 5.21, 4.75, and 2.89 at slow, moderately slow, normal and fast speeds (SDs were 1.38, 1.16, 1.24, and 1.12). A one-way repeated measures ANOVA revealed a significant main effect ($F(3,81) = 23.0$ $p < .001$, $\eta_p^2 = .482$). Multiple comparisons revealed significant differences between slow and moderately slow ($p < .001$), slow and fast ($p < .001$), moderately slow and fast ($p < .001$), and normal and fast ($p < .001$). No significance was found between slow and normal ($p = .333$) or moderately slow and normal ($p = .184$).

Accordingly, for the subjective evaluation of speech rate, the prediction in hypothesis 1b is supported, but the predictions in hypotheses 1a and 1c are not.

Credibility (Figure 2(c)): The scores were 5.14, 5.41, 5.30, and 4.68 at slow, moderately slow, normal, and fast speeds (SDs were 1.15, 1.18, 1.09, and 1.18). A one-way repeated measures ANOVA revealed a significant main effect

(F(3,81) = 5.81, p<.005, η_p^2 = .177). Multiple comparisons revealed significant differences between slow and fast (p = .0167), moderately slow and fast (p = .0390), and normal and fast (p = .0220). Since comparisons among slow, moderately slow, and normal are all far from significant (p > .200), the predictions for credibility in hypotheses 1, 2, and 3 are supported.

Competence (Figure 2(d)): The scores were 4.33, 4.96, 4.52, and 3.71 in slow, moderately slow, normal and fast speeds (SDs were 1.33, 1.19, 1.11, and 1.22). A one-way repeated measures ANOVA revealed a significant main effect (F(3,81) = 10.8, p < .001, η_p^2 = .285). Multiple comparisons revealed significant differences between slow and moderately slow (p = .0427), slow and fast (p = .0313), moderately slow and fast (p < .001), and normal and fast (p < .001). The comparison between moderately slow and normal approaches significance (p = .0623), but no significance was found between slow and normal (p = .333).

Therefore, for competence, the prediction in hypothesis 1b is supported, but the predictions in hypotheses 1 and 3 are not.

Overall result: Hypothesis 1a was not supported except for credibility. Hypothesis 1b was supported in all measurements. Hypothesis 1c was not supported in either subjective evaluations or competence.

4. Experiment 2

Experiment 1 does not support hypothesis 1c. Would it be supported if more involved interaction were conducted that would require more cognitive load, thus overriding people's usual preference for normal to fast speech? As a pragmatic example that would require extra cognitive load, we set a situation where participants walked with the robot, which is a likely scenario of future human-robot interaction. Gait is not a simple automatic motor process, but a demanding cognitive process (Kelly et al. 2008; Al-Yahya et al. 2009; Verrel et al. 2009). Walking decreases cognitive performance if the cognitive task is sufficiently difficult (Srygley et al. 2009).

In this setting, we investigated which speech rate is more appropriate when the robot is walking. The same procedure was used as in Experiment 1, except that the participant and the robot walked together.

A. Conditions

This experiment also used one factor, speech rate, with four conditions: fast, normal, moderately slow, and slow. The details of the conditions are identical to those of Experiment 1 (Section 3.B).

B. Method

Except for the arrangements made for walking, the same method was used as that in Experiment 1; therefore, here we simply explain the parts of the method unique to Experiment 2.

1. *Participants*

Forty-nine native Japanese speakers (25 males, 24 females, average age: 26.6) were paid for their involvement.

2. *Setting*

The experiment was conducted within an area of our research institute that is currently not used for daily business. It includes a couple of vacant rooms connected by a corridor through which people rarely pass. Figure 3 shows the environment and the walking path.

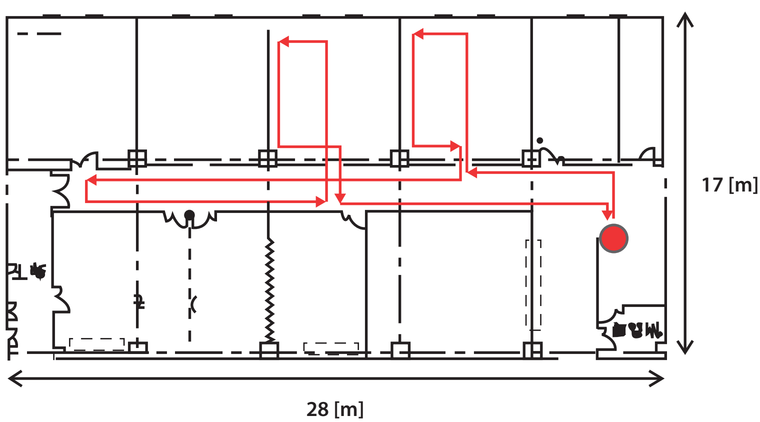


Figure 3. Experimental environment and walking path

When the robot provided speech, it was walking with the participant, who walked around the environment on a pre-defined path (red line in Figure 3); the robot followed the participant's lead. A session was terminated when the robot finished its speech. While walking, the participant listened to the robot (Figure 1(b)). The robot, which maintained gaze directed at the participants, tended to walk slightly behind the person, since its maximum walking speed was set to 450 mm/sec for safety, which is slower than typical human walking speed.

The robot was controlled by an experimenter, who walked about 3 m behind the robot and avoided entering the participant's view.

3. *Procedure*

The procedure in Experiment 1 (Section 3.C.4) was used except that the voice volume was adjusted.

We adjusted the voice volume by asking participants to listen to a short sentence as the robot was walking and adjusted the volume until they were satisfied. This procedure eliminated any possible hearing difficulty caused by a signal-level problem (but not by the cognitive level), such as different distances or influence from the robot's locomotion noise.

4. *Stimulus*

The stimulus from Experiment 1 (Section 3.C.4*)*) was used.

C. Measurements

The measurements from Experiment 1 were used.

- Comprehension: 10-point scale
- Subjective evaluation of speech rate: 1-to-7 (bad to good)
- Competence: 1-to-7 (incompetent to competent)
- Credibility: 1-to-7 (not credible to credible)

D. Hypothesis and predictions

The hypotheses and the predictions made in Section 3. *A* were used.

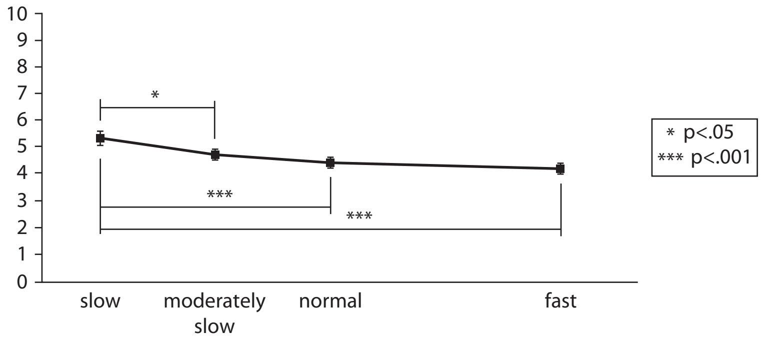
E. Results

Figure 4 shows the results. For multiple comparisons we used the Ryan method (1960).

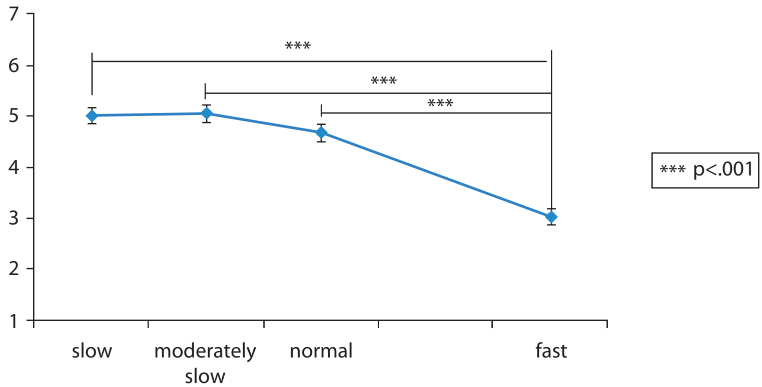
Comprehension: Figure 4(a) illustrates the comprehension results. The scores were 5.22, 4.62, 4.24, and 4.03 at slow, moderately slow, normal and fast speeds (SDs were 1.49, 1.62, 1.90, and 1.66). A one-way repeated measures ANOVA revealed a significant main effect ($F(3,144) = 11.0$, $p < .001$, $\eta_p^2 = .186$). Multiple comparisons revealed significant differences between slow and moderately slow ($p = .0450$), slow and normal ($p < .001$), and slow and fast ($p < .001$). No significance was found between moderately slow and normal ($p = .211$) or normal and fast ($p = .333$). The difference between moderately slow and fast approaches significance ($p = .0980$).

Therefore, for comprehension, the prediction in hypothesis 1c is supported, and the predictions in hypotheses 1 and 2 are not.

Subjective evaluation of speech rate (Figure 4(b)): The scores were 5.01, 5.06, 4.68, and 3.03 at slow, moderately slow, normal and fast speeds (SDs were 1.08,



(a) Comprehension



(b) Subjective evaluation of speech rate

Figure 4. Results from Experiment 2: walking situation *(Continued)*

1.20, 1.20, and 1.13). A one-way repeated measures ANOVA revealed a significant main effect ($F(3,144) = 44.7$, $p < .001$, $\eta^2_p = .482$). Multiple comparisons revealed significant differences between slow and fast ($p < .001$), moderately slow and fast ($p < .001$), and normal and fast ($p < .001$). Other comparisons were far from significant (all $p > .200$). Thus, for the subjective evaluation of the speech rate, the predictions in hypotheses 2 and 3 are supported.

Credibility (Figure 4(c)): The scores were 5.43, 5.73, 5.44, and 4.96 at slow, moderately slow, normal and fast speeds (SDs were 1.12, 0.99, 1.13, and 1.08). A one-way repeated measures ANOVA revealed a significant main effect ($F(3,144) = 9.93$, $p < .001$, $\eta^2_p = .171$). Multiple comparisons revealed significant differences between slow and fast ($p < .001$), moderately slow and fast ($p < .001$), normal and fast ($p < .001$), and moderately slow and normal ($p = .0140$). There is a trend showing a difference between slow and moderately slow, which is approaching significance ($p = .0907$); slow and normal were not significant ($p = .333$).

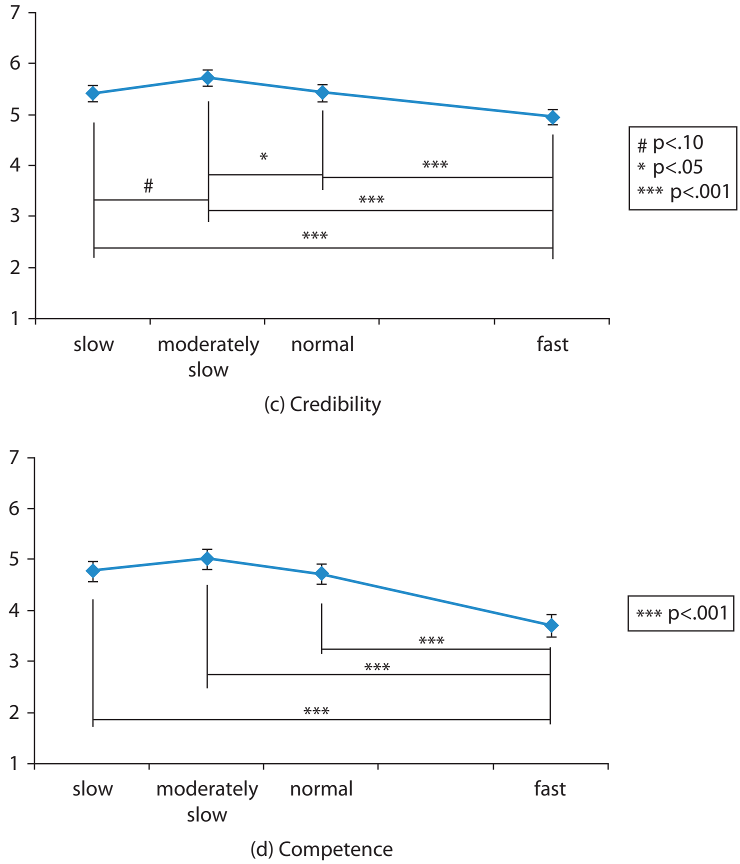


Figure 4. Results from Experiment 2: walking situation

Thus, for credibility, the prediction in hypothesis 1b is supported. The prediction in hypothesis 1c was not counter-proved, but since the comparison between slow and moderately slow was approaching significance, this support is noted with a large caution.

Competence (Figure 4(d)): The scores were 4.79, 5.01, 4.71, and 3.71 at slow, moderately slow, normal and fast speeds (SDs were 1.37, 1.36, 1.44, and 1.57). A one-way repeated measures ANOVA revealed a significant main effect ($F(3,144) = 19.7$, $p < .001$, $\eta_p^2 = .291$). Multiple comparisons revealed significant differences between slow and fast ($p < .001$), moderately slow and fast ($p < .001$), and normal and fast ($p < .001$). Other comparisons were not significant: slow and moderately slow ($p = .333$), slow and normal ($p = .333$), and moderately slow and normal ($p = .157$). Thus, for competence, the predictions in hypotheses 2 and 3 are supported.

Overall results: Hypothesis 1a was not supported at all. Hypothesis 1b was supported except for comprehension. Hypothesis 1c was supported in all measurements.

5. Discussion

A. Summary of findings

Table 1 summarizes the results of the findings related to the three hypotheses. Hypothesis 1a, which was developed based on previous literature in human communication, received little support. In contrast, hypothesis 1b was supported in the standing situation, suggesting that people prefer moderately slow speech; no counter-proof was found for this. In the walking situation, hypothesis 1b was also partially supported. Hypothesis 1c, which was not well supported in the standing situation, is mostly supported in the walking situation. From these results, we give the following interpretations toward answering the initial question we raised: what is the appropriate speech rate for robots?:

1. A robot should speak slower than normal in information-providing tasks.
2. A robot should adjust its speech rate based on context; for example, when walking, people prefer even slower speech.

Table 1. Summary of results related to hypotheses

		Comprehension	Subjective evaluation	Credibility	Competence
Standing	Hypothesis 1a			Supported	
	Hypothesis 1b	Supported*	Supported	Supported	Supported
	Hypothesis 1c	Supported		Supported	
Walking	Hypothesis 1a				
	Hypothesis 1b		Supported	Supported	Supported
	Hypothesis 1c	Supported	Supported	Supported*	Supported

*: with a caution.

B. Comparison of two experiments

In addition to the results shown in each experiment section, we compared the results from Experiments 1 and 2 by running a two-way mixed-design ANOVA for all four measurements. The analysis showed a significant main effect of speech rate on comprehension, $F(3,255) = 12.9$, $p < .001$, $\eta_p^2 = .147$. The conversational scenario (walking or standing) also had a significant effect on comprehension, $F(1,75) = 16.8$, $p < .001$, $\eta_p^2 = .183$. The result showed no significant interaction between speech rate and conversational scenario over comprehension, $F(3,225) = 0.367$, $p = .777$, $\eta_p^2 = .005$.

The analysis showed a significant main effect of speech rate on the subjective evaluation of speech rate, $F(3,225) = 60.8$, $p < .001$, $\eta_p^2 = .448$. But the result showed no significant main effect of speech rate on the subjective evaluation of speech rate, $F(1,75) = 0.215$, $p = .644$, $\eta_p^2 = .003$, and no significant interaction between the speech rate and the conversational scenario over the subjective evaluation of the speech rate, $F(3,225) = 2.25$, $p = .083$, $\eta_p^2 = .029$.

The analysis showed a significant main effect of speech rate on credibility, $F(3,225) = 15.0$, $p < .001$, $\eta_p^2 = .166$. The conversational scenario also had a significant effect on credibility, $F(1,75) = 15.2$, $p < .001$, $\eta_p^2 = .168$. The result showed no significant interaction between the speech rate and the conversational scenario over credibility, $F(3,225) = 1.48$, $p = .222$, $\eta_p^2 = .019$.

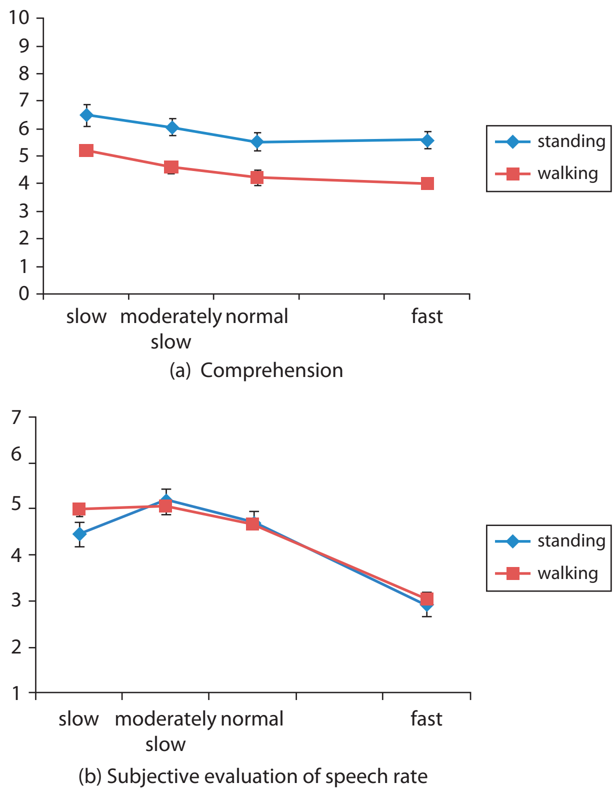


Figure 5. Comparison between results from Experiments 1 and 2 *(Continued)*

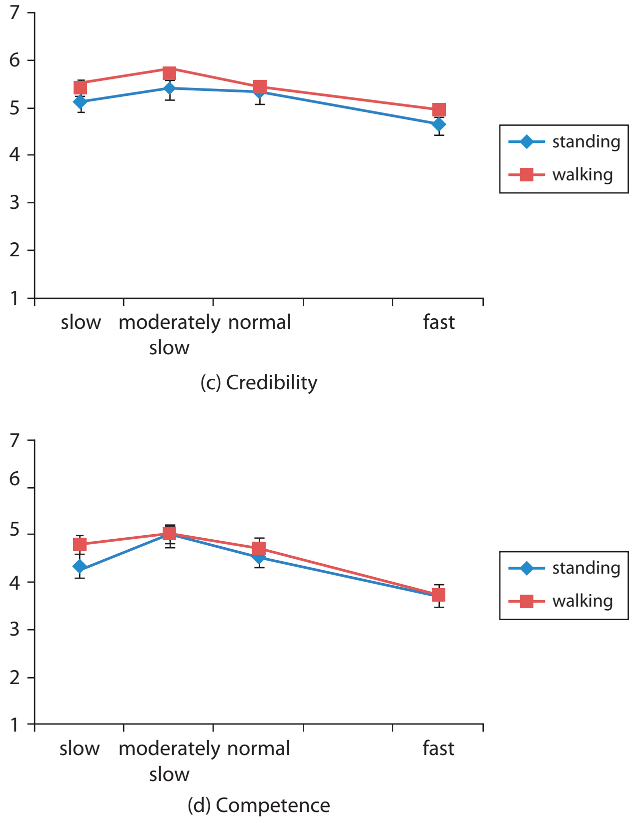


Figure 5. Comparison between results from Experiments 1 and 2

The analysis showed a significant main effect of speech rate on competence, $F(3,225) = 26.3$, $p < .001$, $\eta_p^2 = .260$. The conversational scenario also had a significant effect on competence, $F(1,75) = 39.2$, $p < .001$, $\eta_p^2 < .343$. The result showed no significant interaction between the speech rate and the conversational scenario over competence, $F(3,225) = 1.48$, $p = .222$, $\eta_p^2 = .019$.

In summary, when walking, participants comprehended less but perceived higher credibility and competence. The decrease in comprehension probably reflects the extra cognitive load incurred in walking. In considering the reason for higher credibility and competence, we note that in the walking situation, the robot continuously accompanied the participant. This capability might have led to higher impressions of competence and credibility.

C. Design implications

This study provides the following design implications. First, it confirmed that speech rate is a valid concern for HRI. Even though many researchers probably knew this empirically, this study provides systematic evidence that it affects people's comprehension of speech, their opinion of the appropriateness of the speech rate, and their impression of the robot's capability.

Second, the study provides further evidence that context affects the preferred speech rate. Robots will eventually interact with people in many different contexts. In some cases, people might be focused on the interaction with robots, while in other cases they might be occupied with other activities. Even simple contexts, i.e. walking or standing, are influential. Consequently, when preparing robot utterances, their speed must appropriately reflect the context.

For two scenes, we need to consider using a different speech rate. One immediate application of this knowledge is for guide robots, which have been considered for museums and shops. In a guide scenario, such social robots talk while walking. Our study suggests that robots might need to switch their talking speed depending on whether the context is talking while standing or walking with a person.

Moreover, in an environment where a user might have a higher cognitive load, a slow speech rate is preferred. For example, when a social robot meets a visitor for the first time in a noisy shopping mall, a user might be overwhelmed and perceive a higher cognitive load. In reality, when we put a robot in a shopping mall (Kanda et al. 2010), we received a number of requests for making its utterance slower. We expect the cognitive load to be higher when a robot talks about information that is very new to users, when a user is engaged in a complex task (e.g. assembling parts by following difficult instructions), and when a user is lost. On the other hand, when users probably know what the robot is talking about (e.g. giving casual information or confirming common ground) and when users are not busy with concurrent tasks, they might get frustrated if a robot speaks slowly.

D. Interpretations

1. *Why do people prefer slower speech?*

The experimental results show that people prefer slower speech to fast or normal speech. This contradicts the literature which argued that faster speech was perceived as competent (Buller & Aune 1992) and that speech at normal speed was perceived as credible (Simonds et al. 2006). Perhaps the cause is the synthesized voice, since these works addressed natural human speech. However, a previous study with a synthesized voice reported that fast speech is better for word identification (Roring et al. 2007).

Currently we lack evidence to explain this phenomenon. Our speculation hinges on the presence of extra cognitive load. Tun et al. reported that extra cognitive load decreases sentence recall, and the decrease is larger when speech is fast

(Tun et al. 1992). In our experiment, participants were situated in social interaction with the robot which requires more effort than simply listening. In contrast, previous studies used recorded stimuli, and participants simply listened to a speech from a tape recorder (Buller et al. 1992), video (Simonds et al. 2006), or a computer (Roring et al. 2007). In those experimental trials, there was no social interaction.

2. *Are the phenomena that simple? Alternative interpretations*

In this study, we kept our hypotheses simple because we wanted to choose an appropriate speech rate. Thus, hypotheses 2 and 3 were associated with the measurements of particular speech rates (moderately slow and slow). However, this assumption might be too simple. Our experimental data do not provide evidence to fully distinguish whether our assumption is too simple. Nevertheless, let us briefly discuss the possibility that the underlying phenomena are indeed greatly complex.

Credibility in the walking situation shows ambiguous results for hypothesis 1c without providing significant evidence to counter-prove it. Across conditions, credibility tends to be flat. Related to credibility, in interviews, some participants described the robot as credible because a robot should only speak the truth. This might explain why credibility trends were not sharp across conditions. Note that our questionnaire clearly asks about the credibility of information; however, when participants were asked about credibility, they might be implicitly affected by such a notion that a robot would only speak the truth, and thus were less influenced by its manner of speaking.

In the standing situation, hypothesis 1b's prediction for comprehension was only supported with a caution. Perhaps the robot's slow speech simply provides better comprehension, even though subjectively it is not considered as good as moderately slow or normal speech. Since it remains uncertain whether this alternative explanation is correct, it needs further study.

Note that, in any case, we still need to set a robot's speech rate at a certain level. Even though further studies will reveal more complex relationships among these variables, our study provides effective guidelines to help set a tentatively appropriate speech rate. Perhaps moderately slow speech is preferred in a standing situation and slow speech in a walking situation. With these speeds, we can expect reasonably good comprehension, good subjective evaluations, and favorable impressions of the robot for competence and credibility.

E. Limitations and further studies

The comprehension of spoken language involves the interaction of complex phenomena (Pisoni et al. 1985) and perhaps becomes even more complex when speech is given by a robot. Thus, since a single study cannot cover a large set of parameters, we conducted a systematic study with basic parameters: four speech rates and standing or walking contexts. Apparently, when we use knowledge to control a robot's speech rate, the levels should be adjusted more precisely; perhaps between such levels sharp peaks might be observed if we measured them.

We limited ourselves to standing and walking contexts for cognitive load difference and only tested a single walking speed. The walking settings involve two factors, orientations and people's walking status. While walking, participants were positioned side-by-side in relation to the robot, and while standing, participants were positioned face-to-face, causing different exposure to social cues from the robots. For pragmatic reasons, we did not separate them; yet the findings need to be tested again to consider using a robot in a different combination, for example, standing side-by-side.

Other potential factors include whether the robot is large or small, whether it gestures, and whether it is used in a noisy environment. In addition, we only tested with a single, female adult voice; however, the literature suggests we can only expect a small effect from particular voice-type characteristics (Edworthy et al. 2003), and few studies have reported the interaction between the effect of speech rate and voice characteristics.

The study was conducted with a young population. An elderly population tends to recall less than young people (Tun & Wingfield 1994) and are more distracted by meaningful distracters (Larsby et al. 2005; Tun, O'Kane & Wingfield 2002). Meanwhile, elderly people seem to have a higher capability in reading from context to compensate for lower listening capability (Roring et al. 2007). The relevant literature suggests both directions for conducting our study with an elderly population, who might be less affected by higher cognitive load because they can supplement with their knowledge. Or they might be more affected because it more greatly harms their listening. In either case, the robot needs to carefully adjust its speech rate for an appropriate level of cognitive load.

This study was conducted in Japanese with Japanese people from a Japanese culture. The impression provided by fast speech might be the opposite in English. For example, previous English studies (Buller et al. 1992; Simonds et al. 2006) suggest that perceived intelligence is associated with fast speech, while in Japanese, a study (Tasaki 2003) recommends avoiding fast speech for persuasion. Consequently, what we found in Japanese speech given by a robot might not hold for other languages. This is not what a single study can test, so we believe that the evidence obtained in one language can be useful for initiating similar research in other cultures. Japanese people have a positive image of robots in general, so their impression toward a particular robot might partly reflect such innate cultural expectations. Although we do not expect this aspect to greatly affect the differences in impressions among conditions, it needs to be considered, if in the future such trends do not hold in other cultures, particularly where a robot has the opposite image in general.

In future studies, these effects must be clarified so that we can robustly control the speech rate of robots to fit a particular context. Large individual differences might still be found, since, for example, we already know that the young and the elderly have different listening abilities (Tun et al. 2002).

6. Conclusion

We conducted an experiment on four speech rates: fast, normal, moderately slow, and slow. Although it has been reported in human communication literature that normal or fast speed is most useful, our experimental results with a robot do not show a similar trend. People preferred normal or moderately slow speech. Moreover, when participants listened to the robot while walking, normal speech was less comprehensible. We found that slow speech is considered good in subjective evaluations as well as more effective for comprehension.

Acknowledgement

This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research No. 21680022 and 21118008.

References

Al-Yahya, E., Dawes, H., Collett, J., Howells, K., Izadi, H., Wade, D.T., & Cockburn, J. (2009). Gait adaptations to simultaneous cognitive and mechanical constraints. *Experimental Brain Research*, *199*(1), 39–48.

Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3–4), 255–272.

Breazeal, C., Kidd, C.D., Thomaz, A.L., Hoffman, G., & Berlin, M. (2005). Effect of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2005), pp. 383–388.

Brooks, A.G., & Breazeal, C. (2006). *Working with robots and objects: Revisiting deictic reference for achieving spatial common ground*. Paper presented at the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2006).

Buller, D.B., & Aune, R.K. (1992). The effects of speech rate similarity on compliance: Application of communication accommodation theory. *Western Journal of Communication*, *56*, 37–53.

Buller, D.B., Poire, B.A.L., Aune, R.K., & Eloy, S.V. (1992). Social perceptions as mediators of the effect of speech rate similarity on compliance. *Human Communication Research*, *19*(2), 286–311.

Edworthy, J., Hellier, E., & Rivers, J. (2003). The use of male or female voices in warning systems: A question of acoustics. *Noise and Health*, *6*(21), 39–50.

Endrass, B., Rehm, M., André, E., & Nakano, Y.I. (2008). *Talk is silver, silence is golden: A cross cultural study on the usage of pauses in speech*. Paper presented at the IUI workshop on Enculturating Conversational Interfaces.

Fulford, C.P. (2001). A model of cognitive speed. *International Journal of Instructional Media*, *28*(1), 31–41.

Fulford, C.P., & Zhang, S. (1993). Perceptions of interaction: The critical predictor in distance education. *American Journal of Distance Education*, *7*(3), 8–21.

Goodman, M.A., Robinson, D.H., Robinson, S.L., Skinner, C.H., & Sterling, H.E. (1999). Effects of advertisement speech rates on feature recognition, and product and speaker rating. *International Journal of Listening*, *13*, 97–110.

Gross, H.M., Boehme, H., Schroeter, C., Muller, S., Koenig, A., Einhorn, E., Martin, C., Merten, M., & Bley, A. (2009). Interactive shipping Guide Robots in Everday Use-Final Implementation and Experiences from Long-Term Field Trials, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2009), pp. 2005–2012.

Iwase, T., & Ward, N. (1998). *Pacing spoken directions to suit the listener*. Paper presented at the International Conference on Spoken Language Processing (ICSLP1998).

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.

Johnstone, T., & Scherer, K.R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotion* (pp. 220–235). New York: Guilford.

Jones, C., Berry, L., & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language*, *21*, 641–651.

Kanda, T., Ishiguro, H., Imai, M., & Ono, T. (2004). Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, *92*(11), 1839–1850.

Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., & Hagita, N. (2010). A communication robot in a shopping mall. *IEEE Transactions on Robotics*, *26*(5), 897–913.

Kawai, H., Toda, T., Ni, J., Tsuzaki, M., & Tokuda, K. (2004). *XIMERA: A new TTS from ATR based on corpus-based technologies*. Paper presented at the ISCA Speech Synthesis Workshop.

Kelly, V.E., Schrager, M.A., Price, R., Ferrucci, L., & Shumway-Cook, A. (2008). Age-associated effects of a concurrent cognitive task on gait speed and stability during narrow-base walking. *Journal of Gerontology: Medical Sciences*, *63A*(12), 1329–1334.

Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., & Mitsuishi, M. (2000). *GestureMan: A mobile robot that embodies a remote instructor's actions*. Paper presented at the ACM Conference on Computer-supported cooperative work (CSCW2000).

Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Yamazaki, A. Kuno, Y. Luff, P. & Heath, C. (2008). *Effect of restarts and pauses on achieving a state of mutual orientation between a human and a robot*. Paper presented at the ACM conference on Computer supported cooperative work.

Kuzuoka, H., Suzuki, Y., Yamashita, J., & Yamazaki, K. (2010). Reconfigure Spatial Formation Arrangement by Robot Body Orientation, ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010), pp. 285–292.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*(3), 131–143.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). *A storytelling robot: Modeling and evaluation of human-like gaze behavior*. Paper presented at the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids'06).

Nagaoka, C., Komori, M., Draguna, M., Kawase, S., Yuki, M., Kataoka, T. & Nakamura, T. (2003). *Mutual congruence of vocal behavior in cooperative dialogues; Comparison between receptive and assertive dialogues*. Paper presented at the Human Interface Symposium (in Japanese).

Nixon, C., Anderson, T., Morris, L., McCavitt, A., McKinley, R., Yeager, D. & McDaniel, M. (1998). Female voice communications in high level aircraft cockpit noises-Part II: Vocoder and automatic speech recognition systems. *Aviation, Space, and Environmental Medicine*, *69*(11), 1087–1094.

Okuno, Y., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2009). *Providing route directions: Design of robot's utterance, gesture, and timing*. Paper presented at the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009).

Paul, S. (1992). Pictorial redundancy, processing time, number of propositions, and comprehension of compressed speech. *International Journal of Instructional Media*, *19*(1), 71–78.

Pisoni, D.B., Nusbaum, H.C., & Greene, B.G. (1985). Perception of synthetic speech generated by rule. *Proceedings of the IEEE*, *73*(11), 1665–1676.

Raux, A., & Eskenazi, M. (2009). *A finite-state turn-taking model for spoken dialog systems*. Paper presented at the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL.

Rich, C., Ponsler, B., Holroyd, A., & Sidner, C.L. (2010). *Recognizing engagement in human-robot interaction*. Paper presented at the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010).

Robins, B., Dautenhahn, K., Boekhorst, R.T., & Nehaniv, C.L. (2008). *Behaviour delay and robot expressiveness in child-robot interactions: A user study on interaction kinesics*. Paper presented at the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008).

Roring, R.W., Hines, F.G., & Charness, N. (2007). Age differences in identifying words in synthetic speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(1), 25–31.

Ryan, T.A. (1960). Significance tests for multiple comparison of proportions, variance, and other statistics. *Psychological Bulletin*, *57*(4), 318–328.

Satake, S., Kanda, T., Glas, D.F., Imai, M., Ishiguro, H., & Hagita, N. (2009). How to Approach Humans?: Strategies for Social Robots to Initiate Interaction, ACM/IEEE Int. Conf. on Human Robot Interaction (HRI2009), pp. 109–116.

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). *The utility of affect expression in natural language interactions in joint human-robot tasks*. Paper presented at the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2006).

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2009). How quickly should a communication robot respond? Delaying strategies and habituation effects. *International Journal of Social Robotics*, *1*(2), 141–155.

Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, *166*(1–2), 140–164.

Siegwart, R., Arras, K.O., Bouabdallah, S., Burnier, D., Froidevaus, G., Greppin, X., Jensen, B., Lorotte, A., Mayor, L., Meisser, M., Phillippsen, R., Piguet, R., Ramel, G., Terrien, G., & Tomatis, N., (2003). Robox at Expo.02: A large scale installation of personal robots. *Robotics and Autonomous Systems*, *42*, 203–222.

Simonds, B.K., Meyer, K.R., Quinlan, M.M., & Hunt, S.K. (2006). Effects of instructor speech rate on student affective learning, recall, and perceptions of nonverbal immediacy, credibility, and clarity. *Communication Research Reports*, *23*(3), 187–197.

Srygley, J.M., Mirelman, A., Herman, T., Giladi, N., & Hausdorff, J.M. (2008). When does walking alter thinking? Age and task associated findings. *Brain Research*, *1253*, 92–99.

Takamaru, K., Hiroshige, M., Araki, K., & Tochinai, K. (2000). *A proposal of a model to extract japanese voluntary speech rate control*. Paper presented at the International Conference on Spoken Language Processing (ICSLP2000).

Tasaki, T. (2003). *Zishin-wo-motte-hanasu Nihongo-Koe-ni-dasu Nihongo no waza (Techniques to speak Japanese: How to speak confidently)*. Bungeisha.

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J., & Schulz, D. (1999). Minerva: A Second-Generation Museum Tour-Guide Robot, IEEE Int. Conf. on Robotics and Automation (ICRA1999), pp. 1999–2005.

Tun, P.A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453–467.

Tun, P.A., Wingfiel, A., Stin, E.A.L., & Mecsas, C. (1992). Rapid speech processing and divided attention: Processing rate versus processing resources as an explanation of age effect. *Psychology and Aging*, *7*(4), 546–550.

Tun, P.A. & Wingfield, A. (1994). Speech recall under heavy load conditions: Age, predictability, and limits on dual-task interference. *Aging, Neuropsychology, and Cognition*, *1*(1), 29–44.

Venkatagiri, H.S. (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication*, *10*(2), 96–104.

Verrel, J., Lövdén, M., Schellenbach, M., & Lindenberger, S.S.A.U. (2009). Interacting effects of cognitive load and adult age on the regularity of whole-body motion during treadmill walking. *Psychology and Aging*, *24*(1), 75–81.

Ward, N., & Nakagawa, S. (2002). *Automatic user-adaptive speaking rate selection for information delivery*. Paper presented at the Proc. International Conference on Spoken Language Processing.

Ward, N.G., & Mamidipally, S.K. (2008). *Factors affecting speaking-rate adaptation in task-oriented dialogs*. Paper presented at the Speech Prosody 2008.

Watanabe, T. (1990). The adaptation of machine conversational speed to speaker utterance speed in human-machine communication. *IEEE Trans. on Systems, Man, and Cybernetics*, *20*(2), 502–507.

Woodall, W.G., & Burgoon, J.K. (1983). Talking fast and changing attitudes: A critique and clarification. *Journal of Nonverbal Behavior*, *8*(2), 126–142.

Yamamoto, M., & Watanabe, T. (2006). *Time lag effects of utterance to communicative actions on cg character-human greeting interaction*. Paper presented at the IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2006).

Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2010). A model of proximity control for information – presenting robots. *IEEE Transactions on Robotics*, *26*, 187–195.

Yoshihara, T., Tsutaki, K., & Takahashi, K. (2008). *Speaking rate estimation and utterance analysis of fast speech for high-speed reproduction: A practical example of speech database with speaking rate variations*. Paper presented at the IEICE Technical Report, NLC2008-63 (in Japanese).

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Chichester: John Wiley.

Authors' addresses

Takayuki Kanda
ATR-IRC
2-2-2 Hikaridai, Keihanna Science City
Kyoto 619-0288
Japan

kanda@atr.jp

Michihiro Shimada
ATR-IRC
2-2-2 Hikaridai, Keihanna Science City
Kyoto 619-0288
Japan

m.shimada@atr.jp

Authors' biography

Michihiro Shimada is in Advanced Telecommunications Research Institute International(ATR). He graduated in the Master course in Osaka University, Japan (2006) and graduated in the Bachelor course in Osaka University, Japan (2004). His research interest is cognitive science in robots: human-likeness of robot, human-robot relationship, evaluation of human robot interaction.

Takayuki Kanda received his B. Eng, M. Eng, and Ph.D. degrees in computerscience from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. This author became a Member (M) of IEEE in 2004. From 2000 to 2003, he was an Intern Researcher at ATR Media Information Science Laboratories, and he is currently a Senior Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current researchinterests include intelligent robotics and human-robot interaction.