INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Speech Rate Comparison when Talking to a System and Talking to a Human: A study from a Speech-to-Speech, Machine Translation mediated Map Task

Hayakawa Akira [1], *Carl Vogel* [1], *Saturnino Luz* [2], *Nick Campbell* [1]

[1] School of Computer Science and Statistics, Trinity College Dublin, Ireland
[2] Usher Institute of Population Health Sciences & Informatics, University of Edinburgh, UK

campbeak@tcd.ie, vogel@cs.tcd.ie, S.Luz@ed.ac.uk, nick@tcd.ie

Abstract

This study focuses on the adaptation of subjects in Human-to-Human (H2H) communication in spontaneous dialogues in two different settings. The speech rate of sixteen dialogues from the HCRC Map Task corpus have been analyzed as direct H2H communication, while fifteen dialogues from the ILMT-s2s corpus have been analyzed as a Speech-to-Speech Machine Translation (S2S-MT) mediated H2H communication comparison. The analysis shows that while the mean speech rate of the subjects in the two task oriented corpora differ, in both corpora the role of the subject causes a significant difference in the speech rate with the Information Giver using a slower speech rate than the Information Follower. Also the different settings of the dialogue recordings (with or without eye contact in the HCRC corpus and with or without live video streaming in the ILMT-s2s corpus) only show a negligible difference in the speech rate. However, the gender of the subjects have provided an interesting difference with the female subjects of the ILMT-s2s corpus using a slower speech rate than the male subjects, gender does not show any difference in the HCRC corpus. This indicates that the difference is not from performing the map task, but a result of their adaptation strategy to the S2S-MT system.

Index Terms: speech rate, human-computer interaction, task oriented dialogues

1. Introduction

Speech-to-Speech Machine Translation (S2S-MT) systems are becoming a reality as a way of communication. Microsoft has already released the Skype Translator and the Japanese Ministry of Internal Affairs and Communication has announced that the Tokyo Olympics in 2020 is to use information systems that use multilingual machine mediated communication for 14 languages in Speech-to-Speech (S2S) form, and to achieve this they will first be tested in hospitals, tourist cites and shopping centres.[1]

There has been a lot of research into Human-to-Human (H2H) communication and Human-to-Computer (H2C) communication [1] and Machine Translation (MT) mediated communications, too [2]. However as mentioned by Hara and Iqbel [3], little has been published in Speech-to-Speech Machine Translation (S2S-MT) mediated communication. The Karlsruhe Institute of Technology in Germany is one step closer and uses an S2T-MT system to translate lectures from German into English [4], and following the evaluation of the system from student user feedback, they are now trying to reduce the latency caused due to the characteristics of interpreting [5].

[1] Source: Ministry of Internal Affairs and Communications website (http://www.soumu.go.jp/main_content/000285578.pdf) (last accessed on 2017/06/05).

However we think it is important to look at a simple idea. As Picard [6] and Norman [7] mention, it is important that the design of systems do not frustrate the user. So as not to frustrate the user it is important to understand how the user will use and adapt to a S2S-MT system.

In this study we present preliminary results from the comparison of these two types of communication — H2H and S2S-MT mediated communication — and the difference the subjects of the two corpora had with the speech rate of their utterances.

2. Method

To calculate the deviation in the speech rate, in this study we compared the turn duration of the subject with the duration of the output of the given turn using the TTS system pre-installed in macOS computers.

This method was chosen for this comparison because, for one, it had already been used by the authors [8], and half the data was readily available and also, as mentioned previously by the authors, if the orthodox way of calculating the Words Per Minute (wpm) as $(W/T60)$, where W is the word count per utterance and T is the duration of the utterance was used, the imbalanced nature of the spoken words would create a variance that is difficult to interpret. One example of this is the different wpm value of the turns that are indicated in Table 1. The

Table 1: *Subject utterance duration and wpm sample*

Duration (seconds)	Median	Mean	*SD*
Pebbled shore	1.180	1.132	0.12
Go down	0.926	0.976	0.13
Then where?	0.547	0.554	0.02
Speech Rate (wpm)	Median	Mean	*SD*
Pebbled shore	101.70	106.01	12.12
Go down	129.60	122.97	14.04
Then where?	219.40	216.48	8.80

median wpm values for “Pebbled shore” (an item on the map), “Go down” and “Then where?” are 101.70 wpm, 129.60 wpm and 219.40 wpm, respectively. These wpm values would indicate that on average, one is spoken extremely slowly, the other quite slowly and finally, the last, quite quickly. However there is no way of knowing if the different wpm values of these three multi word turns are the result of a speech rate difference or the difference in the duration to pronounce the given words.

The concept of wpm as a reference of speech rate is already an average of word combinations within the utterance of a minute. The random mixture of words with long and short

http://dx.doi.org/10.21437/Interspeech.2017-1584

duration are mixed into an utterance that is calculated into a quantitive measure. Though the frequency of word length has been previously investigated in written language [9] and also in spoken language [10, 11, 12, 13], given the short utterances of the HCRC Map Task corpus (word count = *Mdn*: 4, *M*: 5.95, *SD*: 6.69) and the ILMT-s2s corpus (word count = *Mdn*: 4, *M*: 5.39, *SD*: 7.34) [14], the low count of the words used in most utterances would not provide an accurate wpm value to provide a reliable quantitive measure.

If the window was a longer, such as the first quarter of the dialogue, there might be enough words to balance out the variance, but not in this situation where the window is a single utterance. The other method of using *syllables per second* as a measure may be more reliable, however this would require highly trained annotators to phonologically segment the data.[2]

Therefore, a reference utterance duration was needed to compare the utterances of the subjects in the two corpora. This reference duration was created by taking the transcription text of all utterances and using the TTS system to read out the text at a speed of 180 wpm. Using PRAAT [16], the TTS system output audio files were segmented using the transcription from the plain text file. Once completed, the start and end times from the segmented files were used to calculate the reference duration of each utterance. This reference utterance duration was then used to calculated a percentage difference with the original subject utterance ($1 - S/T$), where S is the duration of the speaker's utterance and T is the duration of the TTS output, with a positive result indicating speech faster than the ILMT-s2s System TTS output and a negative result indicating slower speech. The resulting values for the examples provided in Table 1 are indicated in Table 2. This method would theoretically remove the variance that is created by the differing duration in pronouncing words of differing lengths, since the reference will also be pronouncing the same word — therefore, creating a more stable reference speech rate. Now with the newly calculated values

Table 2: *List of subject wpm, TTS output wpm and percentage difference of Subject and TTS output speech rates*

Wpm of Subject Utterance	Median	Mean	*SD*
Pebbled shore	101.70	106.01	12.12
Go down	129.60	122.97	14.04
Then where?	219.40	216.48	8.80
Wpm of TTS Reference	**Median**	**Mean**	***SD***
Pebbled shore	151.8	151.9	0.12
Go down	213.4	213.2	0.51
Then where?	253.5	254.1	0.73
Subject / TTS Comparison (%)	**Median**	**Mean**	***SD***
Pebbled shore	−49.50	−43.46	15.39
Go down	−64.39	−73.84	22.59
Then where?	−15.58	−17.37	4.62

indicated under "Subject / TTS Comparison (%)" of Table 2, the previous assumption that the three examples were spoken extremely slowly, quite slowly and quite quickly in the order of "Pebbled shore" – "Go down" – "Then where?", can be reassessed to all examples being slower than the reference TTS output and reordered to "Go down" – "Pebbled shore" – "Then where?".

Hereafter the term "speech rate" will be used to indicate the speech rate difference when compared with the reference 180 wpm TTS output of the same utterance.

3. Material

This study investigates the speech rate of subjects in H2H and S2S-MT communication of task oriented conversation using the Map Task technique. For the H2H communication, sixteen dialogues that used maps 01 and 07 (Figure 1) from the HCRC Map Task corpus [17] were used (§ 3.1), and for the S2S-MT communication data, the fifteen dialogues of the English subjects from the ILMT-s2s corpus [18] were used (§ 3.2).

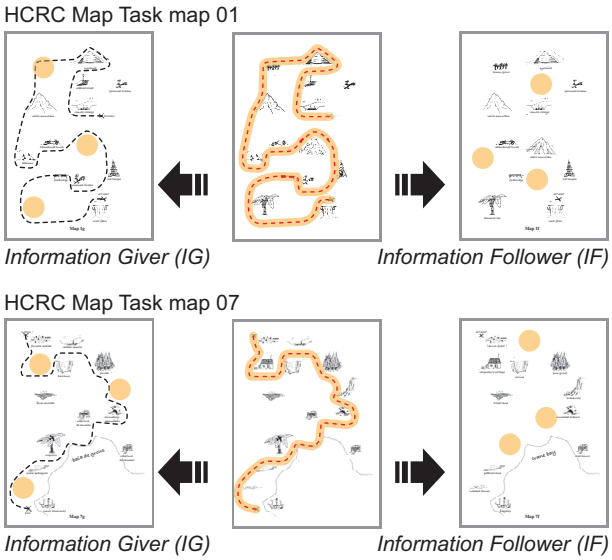

Figure 1: *Map used, with differences highlighted — Left: Map used by IG, Centre: Map with all items, Right: Map used by IF*

3.1. Data from the HCRC Map Task Corpus

The HCRC Map Task corpus contains 128 dialogues of subjects using the map task technique to elicit a dialogue. Among these 128 dialogues, the 16 dialogues that use the same maps as those used in the ILMT-s2s corpus were used in this study.

The dialogues were between native English speakers, mostly from Scotland, half male, half female, half with eye-contact and half without eye-contact. 8 Information Givers (IG) participants (4 ♀, 4 ♂) each performed the role of IG (provide instructions to their interlocutor so they can draw the same route as indicated on their map) twice for the provided map, and 16 Information Followers (IF) participants (8 ♀, 8 ♂) who's role was to replicate the route on their map from the instruction/information provided by the IG.

The dialogue turn segments and text were extracted from the release version 2.1 of the HCRC Map Task corpus data[3] and the extracted segmentation was verified using the dedicated annotation tool ELAN [19] after converting the format for the study of this paper.

[2]The option of using a PRAAT script to automatically calculate the speech rate with the estimator presented by De Jong and Wompe [15] remains a possibility that needs to be explored.

[3]Data downloaded from `http://groups.inf.ed.ac.uk/maptask/maptasknxt.html` (last accessed on 2017/06/05)

3.2. Data from the ILMT-s2s Corpus

The ILMT-s2s corpus contains fifteen dialogues between English and Portuguese subjects speaking to each other in their native language via a Speech-to-Speech (S2S) translation system (ILMT-s2s System). Since this study compares the dialogues of this corpus with the HCRC Map Task corpus, which is only in English, only the dialogues from the English subjects were analysed.

3.2.1. *The ILMT-s2s System*

Two subjects, seated in two different rooms, used the ILMT-s2s System (Figure 2) to communicate with each other. The ILMT-s2s System is a system that uses off-the-shelf components — Automatic Speech Recognition (ASR), Machine Translation (MT) and Text-to-Speech synthesis (TTS) — to perform Speech-To-Speech Machine Translation. It is activated by a "Push-to-talk" button that the subject will click-and-hold for the duration of the utterance and release once the subject has finished. Neither subject can hear the other's voice, since the output of the ASR and MT is provided by a synthetic voice.

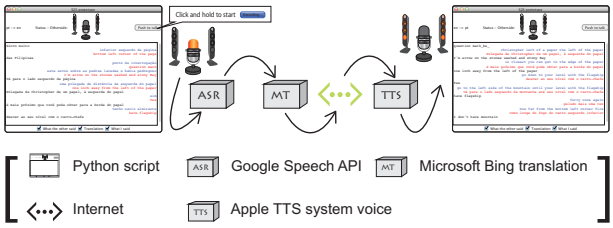

Figure 2: *ILMT-s2s System used to collect the data*

3.2.2. *The Subjects and Recording Environment*

The subjects were recruited from the Trinity College Dublin digital noticeboard or via personal connections. Fifteen recordings of fifteen native English speakers (♀5, ♂10), and fifteen native Portuguese speakers (♀11, ♂4), between the ages of 18 and 45 were collected. Each recording session was conducted in a working office and lasted between 20 and 74 minutes, containing between 33 and 201 On-Talk[4] utterances and between 44 and 212 On-Talk dialogue acts. One subject during each recording session was fitted with biosignals recording device, while the other subject was not (Figure 3).[5]

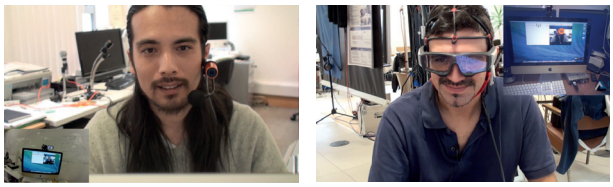

Figure 3: *Subjects during recordings*

3.2.3. *On-Talk, Off-Talk Annotation*

A phenomenon of utterances that are 'not directed to the system' has been described as "Off-Talk" [20]. This phenomenon was also observed in the dialogues of the ILMT-s2s corpus, but instead of studying the type of "Off-Talk" as previously studied [20, 21, 22, 23], here we categorise the talk types by the direction of the utterance. "*On-Talk*"; utterances directed to the interlocutor using the S2S-MT system as a mediator (computer mediated communication), "*Off-Talk Self*"; utterances to oneself, and "*Off-Talk Other*"; utterances directed to a fellow human (direct face-to-face communication).

Since the ILMT-s2s System used a "Push-to-talk" activation method, *On-Talk* locations were retrieved from the system's log file while all other utterances were annotated manually for *Off-Talk Self* and *Off-Talk Other*.

3.3. Summary of the Two Corpora

As mentioned in § 3.1, the HCRC Map Task corpus consists of 128 dialogues, however only 16 dialogues using maps 01 and 07 with a total of 32 subject dialogues, were used in this comparison. As mentioned in § 3.2, the data from the ILMT-s2s corpus consists of 15 dialogues with a total of 30 subjects, however, since we are comparing the data with that of the English speaking HCRC Map Task corpus, the data from the Portuguese counterparts were removed leaving only 15 subjects (IG : IF = 7 : 8). A summary of the turn count in each corpus is listed in Table 3.

Table 3: *Summary of ILMT-s2s corpus (all English dialogues) and HCRC Map Task corpus (all dialogues using maps 1 & 7)*

	All	Single Word	Multi Word
HCRC corpus	3,790	1,393	2,397
ILMT-s2s corpus	1,980	518	1,462
On-Talk	1,328	206	1,122
Off-Talk	652	312	340
Off-Talk Self	483	256	227
Off-Talk Other	169	56	113

A previous study of the ILMT-s2s corpus has shown that the subjects of the ILMT-s2s corpus adapt their speech rate while speaking to the S2S-MT system at a relatively slower speed [8]. Also as differentiated in Table 3, three types of utterances were distinguished within the corpus [24]; *On-Talk* when the subject is using the S2S-MT system as a mediator to communicate with the interlocutor, *Off-Talk Self* when the subject is talking to him/herself, and *Off-Talk Other* when the subject is directly talking to a fellow human. The speech rates of the three talk types of the ILMT-s2s corpus and that of the HCRC Map Task corpus are plotted in Figure 4.

As illustrated in Figure 4, the speech rate box plots of *Off-Talk Other* of the ILMT-s2s corpus and the dialogues of the subjects using maps 01 & 07 of the HCRC Map Task corpus show a similarity. Refer to Table 4 for the median, means and *sd*.

Table 4: *Summary of the various speech rates*

	Median	Mean	*SD*
HCRC corpus	15.65	3.82	43.73
ILMT-s2s On-Talk	−25.54	−35.61	45.24
ILMT-s2s Off-Talk Self	−3.89	−47.43	85.09
ILMT-s2s Off-Talk Other	13.50	3.90	37.51

[4]When the subject is using the ILMT-s2s System to mediate the communication with the interlocutor.

[5]Data from the biosignal recordings were not used in this study.

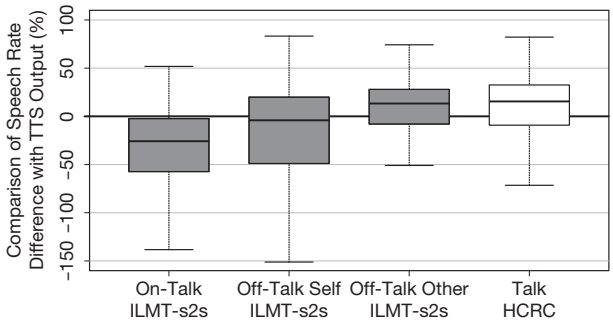

Figure 4: *Speech rates boxplots of both corpora*

Since *Off-Talk Other* is defined as direct communication with a fellow human, the speech rate of *Off-Talk Other* was therefore compared with the speech rate values of the HCRC Map Task corpus. This comparison was made to clarify that there were no significant differences between the two H2H speech rates so as to indicate that the subjects of the ILMT-s2s corpus were not a fluke selection of slower speakers. As a result, no significant difference in the speech rate was observed (Mann-Whitney U test: $p = 0.2565$). Also the effect size was verified for good measure that resulted in a negligible estimate (Cliff's δ estimate: 0.051). This indicates that the non-mediated H2H communication in both corpora use similar speech rates and that the subjects of the ILMT-s2s corpus speak at a similar speech rate in direct H2H communication.

4. Results

The results first study the data of the HCRC Map Task corpus to identify the speech rate patterns of the subject's role, gender, and the recording setting with eye-contact and without eye-contact. Next, the same analysis is performed on the ILMT-s2s corpus data to see if the patterns identified in the HCRC Map Task corpus are followed.

4.1. Analysis of the HCRC Map Task Corpus

A Mann-Whitney U test of the HCRC Map Task corpus groupings of role, gender, and setting indicate that there is a significant difference within the following groupings; Role of subject as Information Giver (IG) or Information Follower (IF) ($p < 2.2e-16$), and Gender of the subject ($p = 0.0093$). However no significant difference was observed for the setting of the recordings (with or without eye-contact (EC)).

Though a significant difference was observed from the groupings of role and gender, the effect size using Cliff's δ estimate shows that only the role difference of Information Giver (IG) or Information Follower (IF) has a reportable difference of "small" — Role: $\delta = 0.300$ (small), Gender: $\delta = 0.061$ (negligible). Boxplots comparing the speech rates with the TTS output duration are plotted in white in Figure 5.

4.2. Analysis of the ILMT-s2s corpus

As with the data of the HCRC Map Task corpus, a Mann-Whitney U test of the ILMT-s2s corpus groupings of role, gender, and setting indicate that there is a significant difference within all the groupings; Role of subject as Information Giver (IG) or Information Follower (IF) ($p = 1.986e-09$), Gender of the subject ($p < 2.2e-16$), and the setting through the availability of eye-contact (EC) with the live video streaming ($p = 0.0019$).

Though a significant difference was observed, the effect size using Cliff's δ estimate shows the following results — Role: $\delta = 0.194$ (small), Gender: $\delta = 0.392$ (medium), and setting (eye-contact): $\delta = 0.099$ (negligible). Boxplots comparing the speech rates with the TTS output duration are plotted in grey in Figure 5.

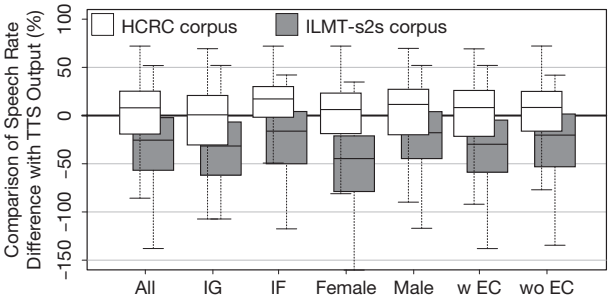

Figure 5: *Speech rate boxplots of both corpora in groupings of role, gender and setting with and without eye-contact*

5. Discussion and Conclusion

The analysis performed in this paper is of extreme simplicity, however the results indicate a gender pattern that are not frequently represented in the literature. Academic literature is split [25, p. 681] on the subject of gender speech rate differences, but the ILMT-s2s corpus data has clearly indicated that within the circumstances of ILMT-s2s corpus data recording, male subjects do not reduce their speech rate as frequently as female subjects, as indicated in Figure 6. A study of disfluency within the

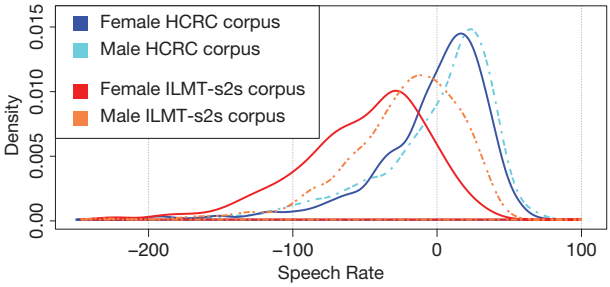

Figure 6: *Density plot of gender speech rates in both corpora*

subjects of the HCRC Map Task corpus [26], also reported that differences in disfluency were found in the role, but not the gender of the subject. Therefore, the gender speech rate difference in ILMT-s2s corpus cannot be explained from the H2H communication of the HCRC Map Task corpus. The adaptation method used for this S2S-MT mediated communication scenario may be related to the differing depth of engagement with language related activities [27] and should merit further investigation.

6. Acknowledgements

This research is supported by Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre (www.adaptcentre.ie) at Trinity College Dublin.

7. References

[1] H. P. Branigan, M. J. Pickering, J. Pearson, and J. F. McLean, "Linguistic alignment between people and computers," *Journal of Pragmatics*, vol. 42, no. 9, pp. 2355–2368, 2010.

[2] N. Yamashita and T. Ishida, "Effects of Machine Translation on Collaborative Work," in *CSCW '06 — 20th Anniversary Conference on Computer Supported Cooperative Work, November 4–8, Banff, Alberta, Canada, Proceedings*, 2006, pp. 515–523.

[3] K. Hara and S. T. Iqbal, "Effect of Machine Translation in Interlingual Conversation: Lessons from a Formative Study," in *CHI'15 — 33rd Annual ACM Conference on Human Factors in Computing Systems, April 18–23, Seoul, Republic of Korea, Proceedings*, 2015, pp. 3473–3482.

[4] M. Müller, S. Fünfer, S. Stüker, and A. Waibel, "Evaluation of the KIT Lecture Translation System," in *LREC 2016 — Tenth International Conference on Language Resources and Evaluation, May 23–28, Portorož, Slovenia, Proceedings*, 2016, pp. 1856–1861.

[5] J. Niehues, T. S. Nguyen, E. Cho, T.-L. Ha, K. Kilgour, M. Müller, M. Sperber, S. Stüker, and A. Waibel, "Dynamic Transcription for Low-Latency Speech Translation," in *INTERSPEECH 2016 — 17th Annual Conference of the International Speech Communication Association, September 9–13, San Francisco, California, USA, Proceedings*, 2017, pp. 2513–2517.

[6] R. W. Picard, *Affective computing*. Cambridge, Massachusetts, USA: MIT press, 2000.

[7] D. A. Norman, *The design of everyday things*. New York, New York, USA: Basic books, 2002.

[8] A. Hayakawa, L. Cerrato, N. Campbell, and S. Luz, "A Study of Prosodic Alignment in Interlingual Map-Task Dialogues," in *ICPhS XVIII — 18th International Congress of Phonetic Sciences, August 10–14, Glasgow, UK, Proceedings*, 2015, paper 0760.1–5.

[9] G. K. Zipf, "The Meaning-Frequency Relationship of Words," *The Journal of General Psychology*, vol. 33, no. 2, pp. 251–256, 1945.

[10] S. Greenberg, "Speaking in shorthand — A syllable-centric perspective for understanding pronunciation variation," *Speech Communication*, vol. 29, no. 2–4, pp. 159–176, 1999.

[11] A. Batliner, E. Nöth, J. Buckow, R. Huber, V. Warnke, and H. Niemann, "Whence and whither prosody in automatic speech understanding: A case study," in *Prosody 2001 — ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding, October 22–24, Red Bank, New Jersey, USA, Proceedings*, 2001, pp. 23–28.

[12] A. Bell, M. L. Gregory, J. M. Brenier, D. Jurafsky, A. Ikeno, and C. Girand, "Which Predictability Measures Affect Content Word Durations?" in *PMLA 2002 — ISCA Tutorial and Research Workshop (ITRW) on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology, September 14–15, Estes Park, Colorado, USA, Proceedings*, 2002, pp. 1–5.

[13] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, no. 1, pp. 92–111, 2009.

[14] A. Hayakawa, S. Luz, and N. Campbell, "Talking to a System and Talking to a Human: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task," in *INTERSPEECH 2016 — 17th Annual Conference of the International Speech Communication Association, September 9–13, San Francisco, California, USA, Proceedings*, 2017, pp. 1422–1426.

[15] N. H. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, no. 2, pp. 385–390, 2009.

[16] P. Boersma and V. van Heuven, "Speak and unSpeak with PRAAT," *Glot International*, vol. 5, no. 9–10, pp. 341–347, 2001.

[17] A. H. Anderson, M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1991.

[18] A. Hayakawa, S. Luz, L. Cerrato, and N. Campbell, "The ILMT-s2s Corpus — A Multimodal Interlingual Map Task Corpus," in *LREC 2016 — Tenth International Conference on Language Resources and Evaluation, May 23–28, Portorož, Slovenia, Proceedings*, 2016, pp. 605–612.

[19] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a Professional Framework for Multimodality Research," in *LREC 2006 — Fifth International Conference on Language Resources and Evaluation, May 22–28, Genoa, Italy, Proceedings*, 2006, pp. 1556–1559.

[20] D. Oppermann, F. Schiel, S. Steininger, and N. Beringer, "Off-Talk — a Problem for Human-Machine-Interaction?" in *EUROSPEECH 2001 Scandinavia: the 7th European Conference on Speech Communication and Technology and the 2nd INTERSPEECH Event, September 3–7, Aalborg, Denmark, Proceedings*, 2001, pp. 2197–2200.

[21] R. Siepmann, A. Batliner, and D. Oppermann, "Using Prosodic Features to Characterize Off-Talk in Human-Computer Interaction," in *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding, October 22–24, Red Bank, New Jersey, USA, Proceedings*, 2001, paper 27.

[22] A. Batliner, C. Hacker, and E. Nöth, "To talk or not to talk with a computer: On-Talk vs. Off-Talk," in *Workshop on 'How People Talk to Computers, Robots, and Other Artificial Communication Partners', April 21–23, Hansewissenschaftskolleg, Delmenhorst, Germany, Proceedings*. SFB/TR 8 Spatial Cognition Report, 2006, pp. 79–100.

[23] ——, "To talk or not to talk with a computer," *Journal on Multimodal User Interfaces*, vol. 2, no. 3, pp. 171–186, 2009.

[24] A. Hayakawa, F. Haider, S. Luz, L. Cerrato, and N. Campbell, "Talking to a system and oneself: A study from a Speech-to-Speech, Machine Translation mediated Map Task," in *SP8 — Speech Prosody 2016, May 31–June 3, Boston, Massachusetts, USA, Proceedings*, 2016, pp. 776–780.

[25] J. V. Borsel and D. D. Maesschalck, "Speech rate in males, females, and male-to-female transsexuals," *Clinical Linguistics & Phonetics*, vol. 22, no. 9, pp. 679–685, 2008.

[26] H. Branigan, R. Lickley, and D. McKelvie, "Non-Linguistic Influences on Rates of Disfluency in Spontaneous Speech," in *ICPhS XIV — 14th International Congress of Phonetic Sciences, August 1–7, San Francisco, California, USA, Proceedings*, San Francisco, California, USA, 1999, pp. 387–390.

[27] E. Roivainen, "Gender differences in processing speed: A review of recent research," *Learning and Individual Differences*, vol. 21, no. 2, pp. 145–149, 2011.