Comparing Perceptual Local Speech Rate of German and Japanese Speech

Hartmut R. Pfitzinger & *Miyuki Tamashima*

Institute of Phonetics and Speech Communication
University of Munich, Germany
{hpt;miyuki}@phonetik.uni-muenchen.de

Abstract

The effect of language background on the perceptual local speech rate (PLSR) is investigated. 160 short German and Japanese speech stimuli are judged by 40 German and Japanese subjects. Japanese listeners overshoot the local speech rate of German speech by 7.5% on a PLSR scale, and German listeners overshoot the speech rate of Japanese speech by 9.1%.

1. Introduction

Possibly everybody who listens to people talking to each other in an unknown language gains the impression that they are speaking very fast. Is this impression an illusion or is there empirical and theoretical evidence? Until today no study primarily addressed this question.

In 1977 Grosjean [3] presented, among other things, French speech to native listeners or listeners who knew no French. He claimed that the processes involved in judging speech rate are purely acoustic (i.e. the number of syllable peaks per second together with the duration and frequency of pauses) and do not require linguistic decoding [4, p. 201]. However, as speech rate estimates made by English listeners with no knowledge of French were usually higher than those of native French listeners, he acknowledged the need for further investigation.

In 1985 den Os [2] examined, among other factors, the effect of language background of Dutch listeners on speech rate perception of nine Dutch and Italian utterances. She concluded "that when listeners are asked to judge [speech] rate differences, they are very well able to do this independent of their language background." In fact, the nine Dutch utterances sounded faster than the nine Italian utterances, which "does not mean that the Italian language generally cannot give the impression of sounding faster than Dutch" [2, p. 133].

These two somewhat contradictory results need clarification. Therefore, in the present paper, we conduct a fully symmetrical perception experiment, in which two groups with different language backgrounds judge the speech rates of stimuli taken from both languages.

2. Method

We chose German and Japanese language since neither of them forms a phonotactic or prosodic subset of the other language.

Instead of presenting whole sentences we decided to use short speech stimuli consisting of only few syllables in order to avoid speech pauses and to reduce any semantic influence.

Pfitzinger 1999 [5] showed that an optimal stimulus duration for judging local speech rate is ca. 625 ms. Below 625 ms an increasing perceptual overshoot is observable. Above 625 ms the probability of local speech rate changes within the stimulus rises making speech rate judgements increasingly difficult.

2.1. Stimuli

A Japanese phonetician perceptually selected 10 speakers (7 female, 3 male) from a Japanese multi-speaker spontaneous speech database who produced the largest speech rate ranges. From each speaker she cut 8 stimuli with durations of 625 ms and with speech rates spread amongst the speaker's whole range of speech rates giving a total of 80 Japanese stimuli.

80 German stimuli consisting of spontaneous speech produced by 10 speakers (7f, 3m) were selected from a former perception study on speech rate [6, p. 169f.] to enable a cross-study comparison of the perception results and thus a reliability test.

2.2. Apparatus and Procedure

During the perception experiment the subjects had to carry out a computer-aided interactive discrimination test using a desktop metaphor on which they could place and reorganize the labels of the 160 speech stimuli and auditorily compare them as often as they wished (see Fig. 1).

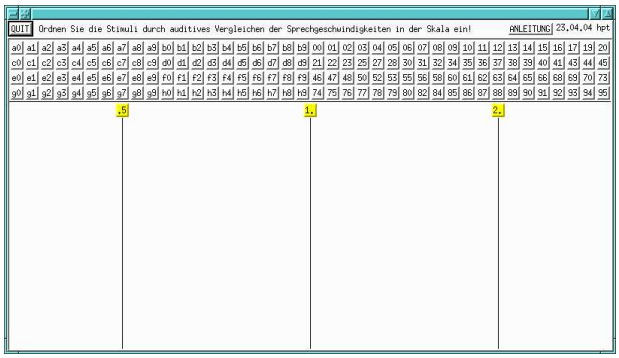

Figure 1: *Graphical user interface of the computer-aided interactive discrimination experiment for judging perceptual local speech rate (PLSR).*

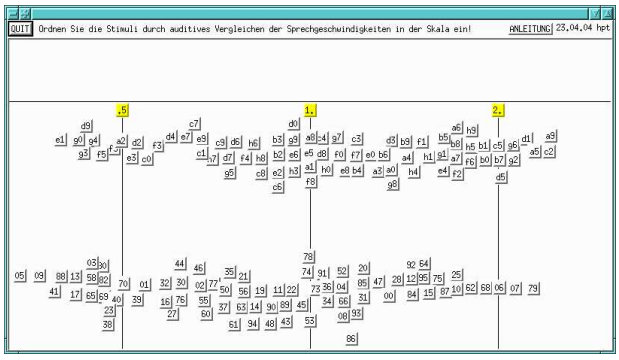

Figure 2: *An example for a finished PLSR judgement task.*

The subjects were instructed to arrange all stimuli along a rate-scale according to the speech rate and to finally check all labels for their correct order, and all perceptual speech rate differences between them for corresponding distances on the rate-scale.

Three anchor stimuli served as a reference for the subjects to orientate to. They were the same as in [5] to guarantee that the subjects would use the desktop space comparably. One of the three anchor stimuli is placed in the middle of the desktop, having a "normal speech rate" which is defined as a typical average speech rate. A stimulus horizontally placed at this position leads to a PLSR value of 100%.

The second anchor stimulus, having roughly half of the normal speech rate, is placed on the left, and the third, having a doubled normal speech rate, is placed on the right. As an example, Fig. 2 shows the final stimulus positions chosen by a particular subject characterized by having spatially divided Japanese and German stimuli.

2.3. Subjects

20 native German and 20 native Japanese subjects with no reported speech or hearing defects took part in the perception experiment. The German subjects were university students who had no knowledge of Japanese. The Japanese subjects were students who were on short vacation in Germany but had no knowledge of German.

3. Results

Four two-way ANOVAs of the raw perception data reveal highly significant factors *stimulus* and *subject* in all four cases (Tables 1–4). This means that both factors influence the variation of perception results which was also the case in [6] where the read speech stimuli explained 72.82% of the total variance and the spontaneous speech stimuli explained 80.81%.

	Deg. of freedom	F	p	Variance explained
Stimulus	79	75.61	0.00	76.94%
Subject	19	15.23	0.00	3.73%
Residual				19.33%

Table 1: *ANOVA of speech rate judgements of German subjects assessing German stimuli.*

	Deg. of freedom	F	p	Variance explained
Stimulus	79	57.48	0.00	72.20%
Subject	19	13.07	0.00	3.95%
Residual				23.85%

Table 2: *ANOVA of speech rate judgements of German subjects assessing Japanese stimuli.*

	Deg. of freedom	F	p	Variance explained
Stimulus	79	64.44	0.00	74.34%
Subject	19	13.48	0.00	3.74%
Residual				21.92%

Table 3: *ANOVA of speech rate judgements of Japanese subjects assessing German stimuli.*

	Deg. of freedom	F	p	Variance explained
Stimulus	79	48.91	0.00	65.91%
Subject	19	26.21	0.00	8.49%
Residual				25.60%

Table 4: *ANOVA of speech rate judgements of Japanese subjects assessing Japanese stimuli.*

Now, the German stimuli explain 77% and 74% of the total variance while the Japanese stimuli only explain 72% and 66% with an increase in unexplained variance of approx. 4% compared with the German stimuli (Tables 1–4). This might be due to the considerably higher background noise of the Japanese stimuli.

A comparison between the variances explained by the factor *subject* reveals that Japanese subjects who judge Japanese stimuli account for 8.5% of the total variance (Table 4) while in the other three cases the factor *subject* explains only 3.73% to 3.95% of the total variance.

A probable reason is that the German language of the three anchor stimuli causes Japanese listeners to accidentally shift, compress, or expand the dispersions of their judgements along the rate-scale and thus decreasing the inter-listener agreements.

On the other hand, Japanese subjects judging German stimuli account for only 3.74% of the total variance which means that they are equally consistent as German subjects judging the same stimuli. Here, the German anchor stimuli do not degrade Japanese assessments.

3.1. Reliability Test

To provide evidence for the reliability of our method we compared the average perception results of the 20 German listeners judging the 80 German stimuli with the average perception results taken from a former study [6] in which 30 German listeners judged the same 80 stimuli.

The scatter plot in Fig. 3 clearly shows that the two groups are strongly correlated. A two-tailed t-test for paired samples revealed that the null hypothesis of equal perception of the two groups could not be rejected even at the 10% level (see Table 5).

This means that the perception results of the former study could reliably be reproduced after four years. Even the presence of Japanese stimuli during the current experiment does not significantly disturb the German subjects' assessments of the German stimuli.

Paired samples	$\hat{t}$	$t_{0.10;79}$	p	
Current vs. former judgements	1.5254	1.6644	0.1312	n.s.

Table 5: *Two-tailed t-test for paired samples applied to current versus former judgements of the 80 German stimuli.*

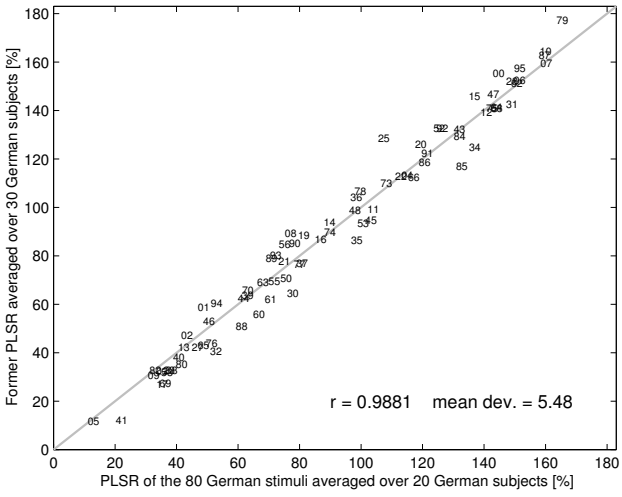

Figure 3: *Perceptual local speech rate scatter plot of judgements of 80 German stimuli averaged over 20 German subjects versus former judgements averaged over 30 subjects.*

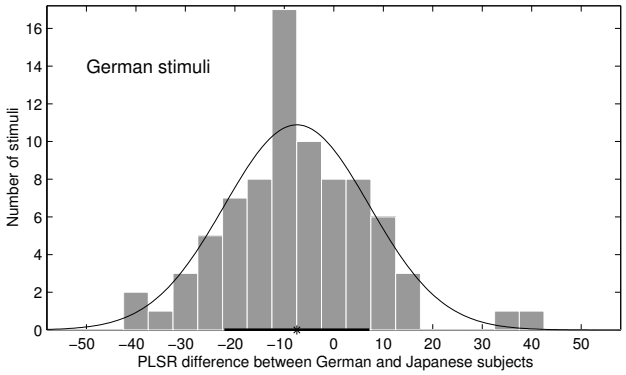

Figure 4: *Difference between average German and Japanese perception of local speech rate of German stimuli.*

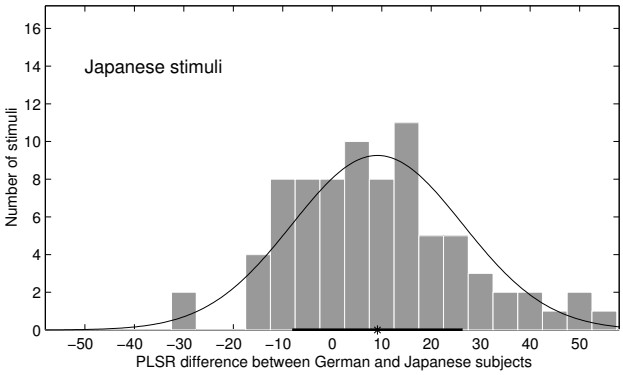

Figure 6: *Difference between average German and Japanese perception of local speech rate of Japanese stimuli.*

3.2. Effect of Language Background

To test the hypothesis that German and Japanese listeners differ in judging speech rate of the same stimuli, two two-tailed t-tests for paired samples were conducted for German and Japanese stimuli, respectively. Table 6 presents the statistical results:

Language of stimuli	$\hat{t}$	$t_{0.001;79}$	p	
German	4.5553	3.4180	0.000019	⋆⋆⋆
Japanese	4.7367	3.4180	0.000001	⋆⋆⋆

Table 6: *Results of two-tailed t-tests for paired samples applied to judgements of German versus Japanese listeners.*

For both languages, the differences between judgements of German and Japanese listeners are highly significant.

The 95% confidence interval of the PLSR difference between German and Japanese subjects judging German stimuli is -10.7 to -4.2. This means that on a perceptual local speech rate scale German stimuli were rated on average 7.47% faster by Japanese subjects than by German subjects. The very opposite is true for Japanese stimuli: They were rated 9.13% faster by German listeners than by Japanese listeners, with a 95% confidence interval of 5.3 to 13.0.

These results are displayed in greater detail in Fig. 4 and 6 which show histograms of the differences between German and Japanese average judgements. First of all, there is no reason to reject the null hypothesis that both histograms are normally distributed, because the total deviances of the German and Japanese distributions are 1.66 and 3.44, respectively, and much smaller than $\chi^2(0.10;5) = 9.24$.

The question arises if the highly significant deviations of -9.13% and 7.47% have different absolute values or are symmetrically spread? Because the alternative hypothesis is that German listeners overestimate Japanese speech rate more than Japanese listeners overestimate German speech rate, a one-tailed t-test is appropriate. The variances of both distributions are homogeneous, $\hat{F} = 1.38 < \mathrm{F}(0.05;79,79) = 1.4512$, n.s.

	$\hat{t}$	$t_{0.10;158}$	p	
German vs. Japanese overshoot	0.6557	1.2869	0.2565	n.s.

Table 7: *One-tailed t-test for independent samples applied to inverted data in Fig. 4 versus original data in Fig. 6.*

As shown in Table 7, there is no significant overestimation difference between German and Japanese subjects. However, Fig. 5 and 7 show that a simple diagonal reflection is not sufficient to match both distributions because German subjects mainly overestimate the speech rate of fast Japanese stimuli while Japanese subjects have a slight tendency to overestimate slow German stimuli more than the fast ones.

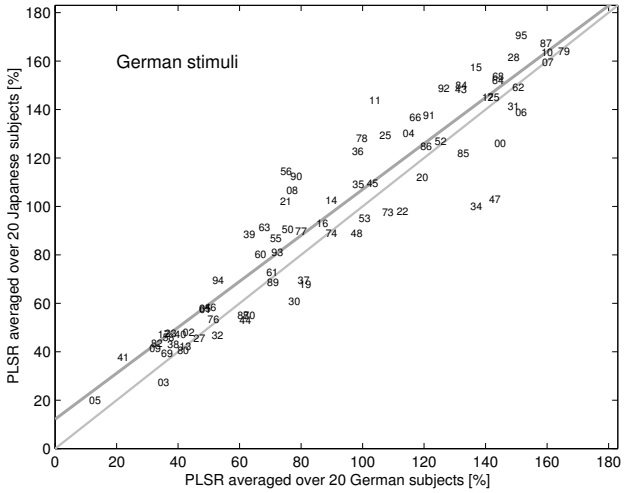

Figure 5: *Scatter plot of German vs. Japanese subjects judging German stimuli. The least squares regression line is dark bold.*

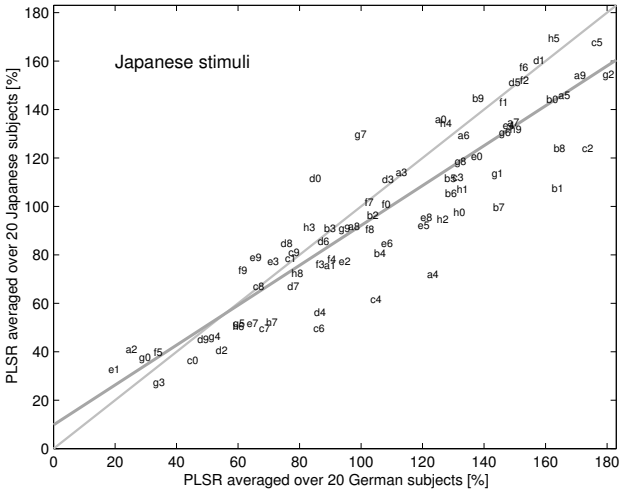

Figure 7: *Scatter plot of German vs. Japanese subjects judging Japanese stimuli. The least squares regression line is dark bold.*

4. Discussion

Our findings clearly support the hypothesis that language background affects the perception of local speech rate. In addition, our quantitative analysis indicates that, on a perceptual local speech rate scale, Japanese listeners overshoot the speech rate of German stimuli by 7.5% and German listeners overshoot the speech rate of Japanese stimuli by 9.1%.

This is in accordance with the outcome Grosjean [3] reported as a by-product of his study, that English listeners with no knowledge of French overestimate the speech rate of French. Even though Grosjean did not perform fully symmetrical perception tests we find evidence from four languages, inspiring us to formulate the more basic hypothesis that a listener with no knowledge of some language overestimates its speech rate in comparison to a native speaker.

Below, we present two possible explanations for this phenomenon, both of which contradict Grosjean's claim [3, 4] and den Os' assumption [2] that linguistic processing is not involved in speech rate judgements. In fact, we suppose that the major part of the unexplained variance of the raw perception data (see Tables 1–4) is due to the variability of the linguistic structure of our stimuli.

4.1. Phonotactic Completion Approach

The inevitable cognitive speech processing generally uses stored knowledge of articulatory and perceptual categories of the native language and thus affects the perception of phones, syllables, and prosodic structure of any unknown language [8].

Usually, a sequence of phonetic items of an unknown language is phonotactically incompatible with the native language. Therefore, unconscious cognitive speech processes insert additional phonetic items into the sequence to reduce the mismatch between its phonotactic structure and the expected structure.

A well-known example of this phenomenon is that Japanese subjects insert reduced vowels into complex consonant clusters of English or German, and in this way create new syllables. Another example is that the unconscious process of reconstructing those phonetic elisions which are common and frequent in the native language, is applied to the unknown language and leads to insertions of new phonetic items.

As a result, the original phonetic structure is perceptually enriched with new phones and even syllables. Consequently, higher phone and syllable rates are perceived which contribute to the impression of a higher speech rate.

4.2. Attenuation and Selection Inability Approach

Another possible approach to explain the overestimation of the rate of unknown speech is based on Broadbent's influential filter theory of attention [1]. It states that due to the limited capacity of mental processing a considerable data reduction of the speech signal is necessary. This is performed by early attentional selection of relevant information. Subsequent theories preferred late selection or attenuation of unheeded information or combinations of these components.

However, when drawing attention to utterances spoken in an unknown language, selection of relevant and attenuation of irrelevant information is virtually impossible due to the lack of phonetic and linguistic knowledge.

Consequently, the attentional focus selects all information available using no specific preference. This leads to a high cognitive load because of the high number of speech items to be processed, and thus to the impression of very fast speech.

5. Conclusion

We tend to combine both approaches: unknown languages appear to be spoken faster because listeners are unable to identify and attenuate redundant features of the unknown speech and, at the same time, they unconsciously insert additional phonetic items to reduce the mismatch between the large number of recognized phonetic items and the phonotactic structure of their native languages.

Finally, our results explain the outcome of den Os [2] that, for Dutch listeners, Dutch sounded faster than Italian: den Os selected stimuli from both languages with pairwise almost identical syllable rates. Consequently, the phone rates of the Italian stimuli were lower than those of the Dutch stimuli because Italian syllables have, on average, fewer phones than Dutch syllables. Since we have shown in 1999 [5, 6] that perceptual local speech rate is strongly correlated with a linear combination of syllable and phone rate, an equal syllable rate in Dutch and Italian combined with a higher phone rate in Dutch should lead to the perception of faster Dutch speech. It seems that the phone rates of her Dutch stimuli were even high enough to mask the effect of perceptual speech rate overshoot for unknown languages, the effect we quantified in the current study.

6. Future Work

A detailed phonetic, phonotactic, and prosodic analysis of those stimuli leading to large inter-group judgement differences, and an acoustic analysis of the Japanese stimuli together with the design of a Japanese PLSR prediction model remain to be done. Although there is strong evidence that a linear combination of syllable and phone rate represents perceptual local speech rate in several languages, this has been shown only for German.

7. Acknowledgements

Many thanks are due to Akiko Nakagawa from Kobe University for making available to us their Japanese multi-speaker spontaneous speech database. We are grateful to Miki Inoue from MPI Munich for her invaluable discussions and inspiration.

8. References

[1] Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press, Oxford.

[2] den Os, E. (1985). Perception of speech rate of Dutch and Italian utterances. *Phonetica*, 42: 124–134.

[3] Grosjean, F. (1977). The perception of rate in spoken and sign languages. *Perception & Psychophysics*, 22(4): 408–413.

[4] Grosjean, F. H.; Lass, N. J. (1977). Some factors affecting the listener's perception of reading rate in English and French. *Language & Speech*, 20: 198–208.

[5] Pfitzinger, H. R. (1999). Local speech rate perception in German speech. In *Proc. of the XIVth Int. Congress of Phonetic Sciences*, vol. 2, pp. 893–896, San Francisco.

[6] Pfitzinger, H. R. (2001). Phonetische Analyse der Sprechgeschwindigkeit. Forschungsberichte (FIPKM) 38, pp. 117–264, Inst. für Phonetik und Sprachliche Kommunikation der Univ. München.

[7] Roach, P. (1998). Some languages are spoken more quickly than others. In Bauer, L.; Trudgill, P., eds., *Language myths*. Penguin Books, Harmondsworth.

[8] Strange, W., ed. (1995). *Speech perception and linguistic experience: Issues in cross-language research*. York Press, Timonium, Maryland.