DOES A TALKER'S OWN RATE OF SPEECH AFFECT HIS/HER PERCEPTION OF OTHERS' SPEECH RATE?

Sandra Schwab
Université de Genève
sandra.schwab@unige.ch

1. ABSTRACT

Many studies have investigated the factors that may affect the perception of speech rate (e.g. Grosjean & Lane, 1976; Feldstein & Bond, 1981; Kohler, 1986; Greene, 1987; Crown & Feldstein, 1991), but very few studies have examined the role that the talker's own rate might play in his/her perception of others' rate. Among them, Lass & Cain (1972) investigated the hypothesis that a speaker's preferred rate depended on his actual rate. They indeed showed that speakers who produced slow rates preferred listening to slow rates, whereas fast speakers tended to prefer fast rates. This conclusion raises the question whether speech rate production affects not only speech rate preference, but also speech rate perception. To our knowledge, very few studies have tried to answer this question. Gósy (1991) formulated the hypothesis that "the speaker's own speech tempo determines his judgements concerning that of other people: the faster his own speech the less fast he perceives that of others" (p. 101). Gósy showed that speakers with different speech rates (very slow, slow, moderate, fast, very fast) did not perceive speech rate in a similar way. In the same direction, Koreman (2006) hypothesized that listeners' own speaking habits may affect their perception of speech rate. Nevertheless, his results failed to show an effect of the listener's rate on his/her perception of rate.

Considering the lack of totally conclusive results on the role that the talker's rate might play in rate perception, the objective of this research is to explore more deeply the hypothesis that speakers with different speech rates do not perceive speech rate in a similar way. To this end, we conducted a perception experiment.

In this experiment, participants were asked to listen to and estimate various samples at different speech rates (normal, fast and slow), using a magnitude-estimation task (Stevens, 1957). Results firstly showed a negative correlation between rate estimation and own rate at normal and slow rates (respectively, $r = -0.45$, $r = -0.39$, $p < 0.05$), but no correlation at fast rate ($r = -0.11$, ns): speakers with fast speech rate tended to under-estimate the sample speech rates (i.e. to give a lower numeric estimation) in comparison with slow speakers (at normal and slow rates). Secondly, and more interestingly, a regression analysis revealed that the own rate has a moderator effect on rate estimation, at all rates (normal: $t(781) = -5.67$, $p < 0.001$; fast: $t(781) = -2.06$, $p < 0.05$; slow: $t(781) = -6.46$, $p < 0.001$): the faster a listener speaks, the less his/her rate estimations raise as a function of heard rates, especially at normal and slow rates.

In sum, the present research has shown that the talker's rate plays a role in rate perception: fast speakers not only tend to under-estimate speech rate in comparison with slow speakers, but they are also less sensitive to rate changes.

2. INTRODUCTION

Speech rate – determined by articulation rate and by number and duration of pauses (see Grosjean & Deschamps, 1975 for a detailed description) – has been widely studied from various points of view for the past 50 years. Among the numerous studies, many have dealt with speech rate perception. The perception of speech rate refers to the metalinguistic activity that performs a listener when hearing a certain rate. In other words, speech rate perception corresponds to the impression a listener gets from the rate of his/her interlocutor. Research in this field has shown that the subjective rate estimation grows more quickly than the objective physical measurements, and that it rises in a non-linear way (Lane & Grosjean, 1973). Speech rate perception can indeed be described by Stevens' power function law (Cartwright & Lass, 1975), which assumes that sensation is proportional to the physical intensity raised to a given power (Stevens, 1957).

Among the researches on factors affecting the perception of speech rate, Grosjean & Lane (1976) showed that articulation rate was more important than pause time in speech rate perception. Acoustic-phonetic factors such as fundamental frequency (e.g. Feldstein & Bond, 1981; den Os, 1985), amplitude (e.g. Feldstein & Bond, 1981) and duration (e.g. Kohler, 1986) have been shown to also affect speech rate perception. Moreover, the influence of cognitive-linguistic variables such as canonical phonological structure (e.g. Koreman, 2006), language (e.g. Grosjean & Lass, 1977), task (e.g. Grosjean, 1978), and language pathology (e.g. Tjaden, 2000) has also been investigated in speech rate perception. Finally, studies have suggested that speech rate perception might vary as a function of extralinguistic factors, such as gender, relationship between speakers (e.g. Crown & Feldstein, 1991) and visual information (e.g. Greene, 1987).

Nevertheless, very few researches have studied the role that the talker's own rate might play in his/her perception of others' rate. For example, Lass & Cain (1972) investigated the hypothesis that a speaker's preferred speech rate depended on his actual speech rate. They showed a good correlation ($r = 0.61$) between speakers' preferred and actual speech rates: speakers who produced slow speech rates preferred listening to slow speech rates, whereas fast speakers tended to prefer fast speech rates. This conclusion raises the question whether speech rate production affects not only rate preference, but also rate perception. To our knowledge, very few studies have tried to answer this question. Gósy (1991) formulated the hypothesis that "the speaker's own speech tempo determines his judgments concerning that of other people: the faster his own speech the less fast he perceives that of others" (p. 101). She indeed showed that speakers with different speech rates (from very slow to very fast) did not perceive rate in a similar way. In the same direction, Koreman (2006) hypothesized that listeners' own speaking habits may affect their perception of speech rate. Nevertheless, his results failed to show an effect of the listener's rate on his rate perception.

Consequently, considering the lack of totally conclusive results on the role that the talker's rate might play in rate perception, the objective of this research is to explore more deeply the hypothesis that speakers with different speech rates do not perceive rate in a similar way. To this end, we conducted the perception experiment that is described below.

3. PERCEPTION EXPERIMENT

3.1 Method

3.1.1 Participants

Twenty-eight French speaking participants took part in this experiment. Their mean age was 27; 9 years.

3.1.2 Stimulus Materials

The stimulus materials used in this experiment consisted of naturally produced versions of a passage at normal, fast and slow rates, recorded by twenty-eight talkers. We used an actualized French version of the "Pop Fan Passage", which has been widely used in studies dealing with speech rate (Grosjean, 1972).

> "A vrai dire, je suis un jeune de quinze ans à peu près normal, ni un cas psychologique sérieux, ni un gars au-dessus des autres. J'écoute Graffiti FM, je coupe mes cheveux très court pour être à la mode, et je porte une boucle d'oreille, mais je ne pense pas être un véritable passionné de musique rap."

Recordings and measurements

Forty native French speakers (20 males and 20 females, mean age of 28 years) were recorded individually in a sound-treated booth. Talkers were instructed to read the passage at rates they considered as normal, fast and slow. Recordings began with three readings at normal rate and continued with three readings at slow rate. After a small break, talkers were asked to read once the passage at normal rate and then three times at fast rate. Each talker's speech was recorded via a microphone onto digital audio tape. As the first reading at each rate and the normal reading after the break served respectively as training and as recalibration, they were not included in the selection procedure.

We measured with *Praat* 3.8 (Boersma, 2001) the duration of speech and pauses for each two readings of each talker. We decided to consider a pause as a silent interval (sometimes with mouth noises and respiration) longer than 200ms, because we wanted to be able to distinguish real pauses from long stop consonant closures, especially at slow rates. Measurements criteria were the following: glottalization before a vowel, aspiration after a stop consonant, release schwa appearing at the end of some consonants and creaky voice were included in speech, whereas aspiration before speech, eventual sighs and mouth noises were included in pauses. From these measurements, we obtained the total speech time, the articulation time, the pause time, as well as the number of pauses.

As far as the syllable number was concerned, we identified and counted syllables in all productions on the basis of listening only. We made sure that our procedure was reliable by asking two judges to identify and count syllables in a subset of productions. As comparisons between judges showed similar syllable number and identification, we obtained speech rate (syll/min), articulation rate (syll/sec), as well as pause number and mean duration (msec) for both readings at each rate for each subject.[1] Then, the mean across the two readings for each variable was computed for each talker, as well as the rate range (difference between slow and fast rates (syll/min)).

Participants and stimuli selection

Although the other temporal variables (articulation rate, pause number and duration) were also considered, the procedure selection was mainly based on the distribution of speech rate (syll/min). We chose a representative subset of productions according to the distributions of normal, fast and slow speech rates. We indeed selected the productions in such a way that their distributions at normal, fast and slow speech rates as well as the distribution of rate range (difference between slow and fast rates) matched as best as possible the entire set of productions. According to this criterion, we selected one production at each rate (normal, fast and slow) of 28 talkers (14 males and 14 females). We made sure that the differences between rate distributions of the entire set ($n = 40$) and rate distributions of the subset of productions ($n = 28$) were not significant (normal: $t(66) = 0.01$, ns.; fast: $t(66) = 0.28$, ns.; slow; $t(66) = 0.7$, ns.), nor was the difference between range significant ($t(66) = 0.3$, ns.).[2] Moreover, despite the overlap between the three speech rates (slow rate of some talkers corresponded sometimes to normal or even fast rate for other talkers, and inversely, fast rate of some talkers was closer to normal rate for others), statistical analyses showed not only a significant rate difference ($F(2, 54) = 243.83$, $p < 0.0001$), but also significant differences between each rate (Tukey HSD, $p < 0.01$).

The selection of the productions enabled us then to invite the 28 talkers who produced them to the perception experiment. In sum, according to the speech rate (normal, fast and slow) and the range distributions, we selected not only materials for the perception experiment – 28 productions at each of the rates (normal, fast and slow) –, but also the 28 talkers – 14 males and 14 females – who would participate in the perception experiment.

3.1.3 Procedure

The 84 selected productions (28 talkers x 3 rates) were split up into three parts (A, B and C), in such a way that one production of each talker appeared in each part, and that no more than two same rates (normal, fast or slow) followed each other. Moreover, we added one filler production at the beginning of each part, and three filler productions were chosen as practice.

Participants were run individually, or two by two. After listening to each production through headphones, they were instructed to perform a magnitude estimation task (see Stevens (1957) for details). In this task, the listener has to assign a number to each speech

[1] The important distinction between speech rate and articulation rate has to be kept in mind. The former refers to the number of units (e.g. words, syllables, phones) produced in a specific time, including pauses. The latter refers to the number of units expressed in a specific time, excluding pauses (Grosjean & Deschamps, 1975). Speech rate can be expressed in syll/min (Grosjean & Deschamps, 1975), in words/min (Goldman-Eisler, 1968) or in phones/sec (Gósy, 1991), while articulation rate is generally expressed in syll/sec (Grosjean & Deschamps, 1975) or in phones/sec (Koreman, 2006).

[2] We also made sure that the other temporal variables (articulation rate, pause number and duration) were similar between the 40 productions and the selected subset of productions. None of the differences was significant ($p > 0.14$).

rate he hears. The number 10 corresponds to what the listener considers a normal speech rate. Thus, the number he gives must be proportional to the normal speech rate (10). For example, 20 corresponds to a speech rate which is twice as fast as the normal speech rate, and 5 corresponds to a speech rate which is twice as slow as the normal speech rate.[3]

Each session consisted of the presentation of the three practice productions, followed by the presentation of the three parts (84 productions), with a break between each part. Half the participants heard the three parts in the order A, B and C, and the other half heard them in the reverse order (C, B and A).

3.1.4 Data analysis

Within each rate (normal, fast and slow) we collected the rate estimation of the 28 heard productions, given by the 28 participants (784 data for each rate). We also computed, within each rate, the mean estimation for each participant (28 data for each rate). By means of correlations and regression analysis, we explored the relationship between rate estimations given by the participants and their own speech rates. Remind that participants' own rates (normal, fast and slow, expressed in syll/min) were obtained thank to the readings they were instructed to do in the recording session described above.

3.2 Results and discussion

This experiment aimed at examining the relationship between speech rate production and perception using a magnitude estimation task. We first connected production and perception data by means of correlations. More precisely, we correlated participants' own rate and their mean rate estimation, separately for the normal, fast and slow rates. Remind that participants read the passage at normal, fast and slow rates and that they had to judge the rate of normal, fast and slow speech samples.

Figure 1 represents rate estimation as a function of own rate for normal, fast and slow rates. Rate estimation is presented in logarithmic values on the left y-axis, while the corresponding raw values are presented in the right y-axis. In the same way, own rate appears in logarithmic values on the lower x-axis, whereas the corresponding values in syll/min appear in the upper x-axis.

[3] Participants could use whatever number they wanted; they were not limited to the numbers 5, 10 and 20.

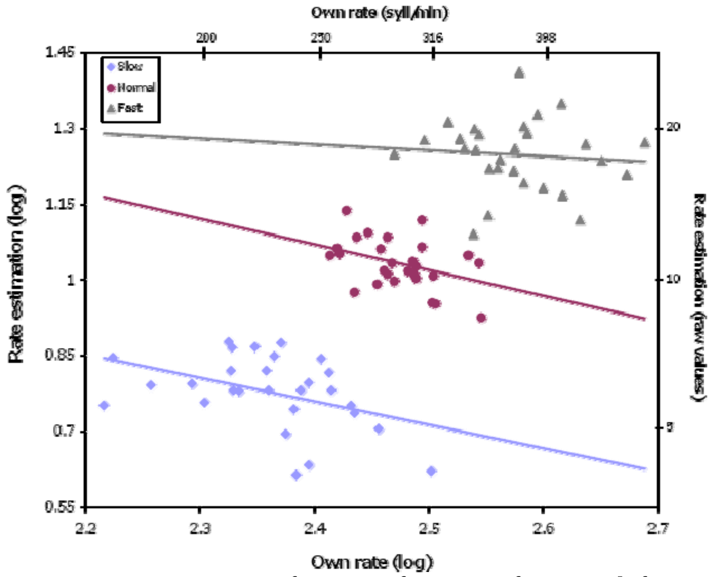

Figure 1: Rate estimation as a function of own rate for normal, fast and slow rates

Listeners are able to distinguish between the three rates: they give estimations that differ significantly between the rates (normal = 10.82, fast = 17.91, slow = 6.07; $F(2, 54) = 488.84$, $p < 0.001$)[4]. Consequently, it seems that the overlap we mentioned between the three rates in production doesn't impair the listeners' rate differentiation ability in the estimation task.

Secondly, and more interestingly, as can be seen in Figure 1, we find a negative correlation between rate estimation and own rate at normal and slow rates (respectively, $r = -0.45$, $r = -0.39$, $p < 0.05$), but no correlation at fast rate ($r = -0.11$, ns). These results show that speakers with fast speech rate tend to under-estimate the sample speech rates (i.e. (to give a lower estimation number) in comparison to slow speakers, especially at normal and slow rates

These results only partly confirm the hypothesis of a relationship between speech rate production and speech rate perception. This relation exists at normal and slow rates but not at fast rate. Further investigation is needed to examine in more details the reasons leading to the absence of a correlation between speech rate production and speech rate perception for fast rates. Indeed, fast speech rate might result in articulatory cues such as deletions (e.g. schwa deletions) or blurred speech, which might facilitate the perception of fast speech (Koreman, 2006), whatever the listener's own rate may be.

Further analyses were conducted in order to determine how the relationship between heard rate and rate estimations varied as a function of own rate. As illustrated in Figure 2, we were interested in studying whether the relationship between Heard rate (Independent

[4] Note that we used logarithmic values in statistic analyses, but for sake of clarity we present means in raw data.

variable, IV) and Rate estimation (Dependent variable, DV) varies according to Own rate, which might play the role of moderator.

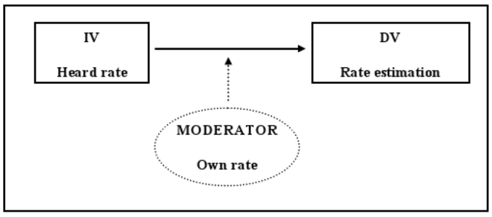

Figure 2: Relationship between Heard rate (IV) and Rate estimation (DV) according to Own rate (Moderator)

Given that moderator variables are characterized statistically in terms of interactions, we included in our regression model the interaction (i.e. cross-product term) of Heard rate and Own rate.[5] Therefore, we ran three separate regression analysis for the normal, fast and slow rates, with Rate estimation as a dependent variable, and Heard Rate, Own Rate and Interaction as independent variables. As the three independent variables (Heard rate, Own rate and Interaction) were highly correlated ($VIF > 10$),[6] the regression analyses were performed with Rate estimation as the dependent variable, and Heard rate and Interaction as independent variables ($VIF = 2$), separately for the normal, fast and slow rates.

As expected, regression analyses show first an effect of Heard rates, at all rates (normal: $t(781) = 21.05$, $p < 0.001$; fast: $t(781) = 19.25$, $p < 0.001$; slow: $t(781) = 24.01$, $p < 0.001$), meaning that Heard rate has a strong impact on Rate estimation. In other words, listeners are able to perceive rate differences within normal, fast and slow speech samples, respectively. Secondly and more interestingly, results show a significant interaction Heard rate x Own rate on Rate estimation, at all rates (normal: $t(781) = -5.67$, $p < 0.001$; fast: $t(781) = -2.06$, $p < 0.05$; slow: $t(781) = -6.46$, $p < 0.001$): the faster a listener speaks, the less his/her rate estimations raise as a function of heard rates.

Figure 3 shows the conditional slope of Rate estimation on Heard Rate as a function of own rate, for normal, fast and slow rates. In other words, the figure shows, on the y-axis, the conditional slopes relating Heard rate and Rate estimation (conditional on Own rate), and on the x-axis, the Own rate (in logarithmic values in the lower x-axis, and in syll/min in the upper x-axis).

[5] Interaction refers here to the cross-product term of Heard rates and Own rate. For example, if Heard rate = 2.54 (in log), and Own rate = 2.46 (in log), thus Interaction = 2.54 x 2.46 = 6.25.
[6] The Variance Inflation Factor (VIF) "provides an index of the amount that the variance of each regression coefficient is increased relative to a situation in which all of the predictor variables are uncorrelated. […] A commonly used rule of thumb is that any VIF of 10 or more provides evidence of serious multicollinearity involving the corresponding IV" (Cohen, Cohen, West & Aiken, 2003; 423).

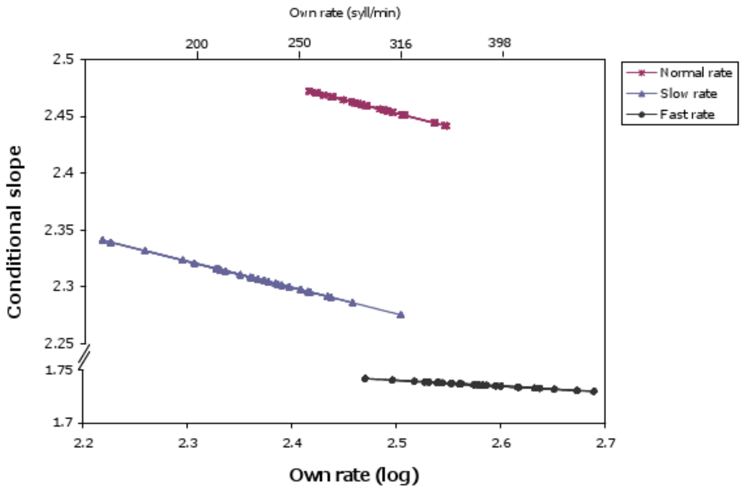

Figure 3: Conditional slope of Rate estimation on Heard Rate as a function of own rate, for normal, fast and slow rates

As can be seen, for all rates, the conditional slope decreases as a function of own rate. We can also observe that the slope is steeper at normal and slow rates than at fast rate, suggesting that own rate has a smaller effect on rate estimation at fast rate. In sum, own rate plays a moderator role in rate estimation: listeners with different speech rates perceive rate in a different way. The faster the own rate is, the less the rate estimation rises as a function of heard rates, especially at normal and slow rates.

4. CONCLUSION

The hypothesis we explored in this research was that speakers with different rates do not perceive speech rate in a similar way. More specifically, we hypothesized that fast speakers tend to under-estimate speech rate (i.e. to give a lower numeric estimation) in comparison with slow speakers. On one hand, participants were asked to read a passage at normal, fast and slow rates, and on the other hand, they were instructed to listen to and estimate various speech samples produced at different speech rates (normal, fast and slow), using a magnitude-estimation task.

Correlation analyses showed that speakers with fast speech rate tend to under-estimate the sample speech rates (i.e. to give a lower numeric estimation) in comparison to slow speakers (at normal and slow rates). Furthermore, regression analyses, which examined the moderator effect of own rate on rate perception, revealed that the faster the own rate is, the less the estimation rises as a function of heard rates. Therefore, the correlations between own rate and estimation, on one hand and, on the other hand, the moderator effect of own rate on rate perception suggest the existence of a relationship between speech rate production and perception, the former defining the latter.

As far as fast rate is concerned, further investigation is needed to study deeper the weakness of the relationship. Indeed, it might be due to a ceiling effect, more specifically to the fact that fast speech is easy to identify in presence of eventual blurred speech or deletions, which would explain why fast speech is perceived as fast by all participants, whatever their own rate may be.

A question that may arise from these results concerns the direction of the relationship between speech rate production and perception. At the segmental level, the direction of the link between production and perception has been considered in both ways. Indeed, following the hypothesis of Perkell *et al.* (2004), speech perception affects speech production, while according to other researchers (Paliwal, Lindsay & Ainsworth, 1983) and defenders of the *Motor Theory of Speech Perception* (Liberman & Mattingly, 1985), speech production regulates speech perception. Following Gósy (1991) and Koreman (2006), we hypothesized that rate production regulates rate perception, but it would be worth considering the reverse possible interpretation.

In sum, we can conclude that a talker's own rate of speech does affect his/her perception of others' speech rate. More specifically, fast speakers not only tend to under-estimate (i.e. to give a lower numeric estimation) speech rate in comparison with slow speakers, but they are also less sensitive to rate changes. This finding highlights the importance of considering and controlling the listeners' own rate in experiments dealing with speech rate perception.

ACKNOWLEDGEMENTS

The author thanks Prof. François Grosjean (University of Neuchâtel, Switzerland) for his help in designing the experiment, and Prof. Boris Wernli (University of Neuchâtel, Switzerland) for his help in statistics.

5. REFERENCES

Boersma, P. (2001), PRAAT, a system for doing phonetics by computer, *Glot International*, 5, 341-345.

Cartwright, L.R. & Lass, N.J. (1975), A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling, *Language and Speech*, 18, 358-365.

Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. (2003), *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences* (Third Edition), Mahwah, NJ: Lawrence Erlbaum Associates.

Crown, C.L. & Feldstein, S. (1991), The perception of speech rate from the sound-silence patterns of monologues, *Journal of Psycholinguistic Research*, 20, 47-63.

den Os, E. (1985), Perception of speech rate of Dutch and Italian utterances, *Phonetica*, 4, 124-134.

Feldstein, S. & Bond, R. (1981), Perception of speech rate as a function of vocal intensity and frequency, *Language and Speech*, 24, 387-395.

Goldman-Eisler, F. (1968), *Psycholinguistics: Experiments in Spontaneous Speech*, London: Academic Press.

Gósy, M. (1991), The perception of tempo, in *Temporal Factors in Speech. A collection of Papers* (M. Gósy, editor), Budapest: Research Institute for Linguistics, HAS, 63-107.

Green, K.P. (1987), The perception of speaking rate using visual information from a talker's face, *Perception & Psychophysics*, 42, 587-593.

Grosjean, F. (1972), *Le rôle joué par trois variables temporelles dans la compréhension orale de l'anglais étudié comme seconde langue, et perception de la vitesse de lecture par des lecteurs et des auditeurs*, Thèse de Doctorat, Université de Paris VII, Paris, France.

Grosjean, F. (1978), *Perception of rate and processing of sentences*, Unpublished Manuscript, Northeastern University, Boston, USA.

Grosjean, F. & Deschamps, A. (1975), Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation, *Phonetica*, 31, 144-184.

Grosjean, F. & Lane, H. (1976), How the listener integrates the components of speaking rate, *Journal of Experimental Psychology: Human Perception and Performance*, 2, 538-543.

Grosjean, F. & Lass, N. (1977), Some factors affecting the perception of reading rate in English and in French, *Language and Speech*, 20, 198-208.

Kohler, K.J. (1986), Parameters of speech rate perception in German words and sentences: Duration, F0 movement, and F0 level, *Language and Speech*, 49, 115-139.

Koreman, J. (2006), Perceived speech rate: The effect of articulation rate and speaking style in spontaneous speech, *Journal of the Acoustical Society of America*, 119, 582-596.

Lane, H. & Grosjean, F. (1973), Perception of reading rate by speakers and listeners, *Journal of Experimental Psychology*, 97, 141-147.

Lass, N.J. & Cain, C.J. (1972), A correlational study of listening rate preferences and listeners' oral reading rates, *Journal of Auditory Research*, 12, 308-312.

Liberman, A.M. & Mattingly, I.G. (1985), The motor theory of speech perception revised, *Cognition*, 21, 1-36.

Paliwal, K.K., Lindsay, D. & Ainsworth, W.A. (1983), Correlation between production and perception of English vowels, *Journal of Phonetics*, 11, 77-83.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. & Zandipour, M. (2004), The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts, *Journal of the Acoustical Society of America*, 116, 2338-2344.

Stevens, S.S. (1957), On the psychophysical law, *Psychological Review*, 64, 153-181.

Tjaden, C. (2000), A preliminary study of factors influencing perception of articulatory rate in Parkinson disease, *Journal of Speech, Language and Hearing Research*, 43, 997-1010.