

Relationship between Speech Rate Perceived and Produced by the Listener

Sandra Schwab

Ecole de Langue et de Civilisation Françaises, Université de Genève, Genève, Switzerland

Abstract

The present research examines the impact of listeners' own rate of speech production on their perception of speech rate. We recorded 28 native French speakers reading a passage at normal, fast and slow rates. The same speakers then assessed the rate of production of these samples by all speakers and at all rates (normal, fast and slow), using a magnitude estimation task. We show that there is an inverse relationship between listeners' own rate of production and their judgment of rate, so that listeners with slower rates tend to overestimate sample rates relative to faster speakers. This effect of listeners' own rate was present in the magnitude estimation results at slow and normal rates, but not at fast rates. Furthermore, listeners with a slow rate are more sensitive to rate differences within a given rate category (normal, fast and slow). The listener's own rate should therefore be considered as a critical factor in speech rate perception.

Copyright © 2012 S. Karger AG, Basel

Introduction

Among the abundant research dealing with speech rate (also called 'speaking rate', 'tempo'), many have focused on its perception, i.e. the sensation of speed a listener gets from the rate of his interlocutor. The multidisciplinary nature of speech rate perception makes its investigation particularly interesting. For example, from a clinical viewpoint, the study of the perception of speech rate in patients with speech disorders (aphasia, Parkinson's disease, etc.) may help to determine the neuroanatomical brain regions that support the perception of temporal information in speech [e.g. Schirmer, 2004]. Speech rate perception can also take on a social dimension. According to Street and Brady [1982] and Putman and Street [1984], the impression a listener gets from the speech rate of a speaker considerably influences the judgments that the former makes about the competences and the social attractiveness of the latter. Finally, speech rate perception can be considered as a communicative marker, as speech rate may be perceived as 'an indicator of the importance of what is being said' [Quené, 2007, p. 353], or as a source of information to be modeled for speech synthesis purposes [e.g. Pfitzinger, 1999].

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
0031–8388/11/0684–0243
\$38.00/0
Accessible online at:
www.karger.com/pho

Dr. Sandra Schwab
Université de Genève, ELCF, Uni Bastions
5, Rue de Candolle
CH–1211 Genève 4 (Switzerland)
Tel. +41 22 379 74 54, E-Mail sandra.schwab@unige.ch

Over the past 40 years, researchers have carried out several studies in order to characterize the perception of speech rate. They have asserted that subjective rate estimate grows more quickly than the objective physical measurements, and thus rises in a nonlinear way [Lane and Grosjean, 1973; Grosjean, 1977]. Speech rate perception can be described by Stevens' [1957, 1975] power function law [Cartwright and Lass, 1975], which assumes that sensation is proportional to the physical measurement raised to a given power. On a logarithmic plot, the power function is usually represented by a straight line the slope of which corresponds to an exponent. According to Lane and Grosjean [1973], who make an important distinction between the perception of autophonic rate (i.e. the perception of a speaker's own rate) and extraphonic rates (i.e. the perception of others' rate), this exponent is 2.6 for the perception of autophonic rate, and 1.5 for the perception of extraphonic rate. In other words, a speaker who doubles his physical rate perceives it as 6 times as fast, whereas the listener perceives it as 3 times as fast. Furthermore, Hoequist and Kohler [1986] and Quené [2007] have shown that listeners are able to detect relatively small rate variations. They showed that the human ear is sensitive to a 20-ms just noticeable difference in syllabic duration, and to a 5% speech rate change.

Moreover, researchers have attempted to determine the factors that affect speech rate perception. As is well known, speech rate is a complex temporal variable based on articulation time as well as on pause time [see Grosjean and Deschamps, 1975, for a detailed description of temporal variables]. Grosjean and Lane [1976] aimed therefore at defining the role of pause time and articulation rate – the speed at which a speaker articulates an utterance, without considering the pauses that might be produced – in the perception of speech rate. They showed that articulation rate was more important than pause time in speech rate perception. Acoustic-phonetic factors have been shown to influence speech rate perception as well. Besides duration, which is obviously crucial in speech rate perception [e.g. Kohler, 1986], fundamental frequency (F0) seems to be a decisive factor [e.g. Feldstein and Bond, 1981; den Os, 1985; Kohler, 1986; Rietveld and Gussenhoven, 1987; Gósy, 1992]. High F0 is a cue for fast rate, while lower F0 characterizes a slower rate [e.g. for German, Kohler, 1986; for Dutch, Rietveld and Gussenhoven, 1987]. Moreover, melodic variations have an impact on rate perception: a rising intonation gives the impression of a faster rate than a falling intonation [e.g. for Hungarian, Gósy, 1992; for German, Kohler, 1986]. Conclusions about the perception of utterances with little melodic F0 variation (monotonous utterances) are not consistent. While a monotonous utterance is perceived as faster than an utterance with melodic variation in Dutch and in Italian [den Os, 1985], it is judged as slower in Hungarian [Gósy, 1992]. Feldstein and Bond [1981] showed that English utterances with high amplitude sound faster than utterances with lower amplitude.

Syllabic structure also plays a role in speech rate perception: thus, a complex syllable (e.g. /plats/) is perceived as faster than a simple syllable (e.g. /pa/) [Hoequist and Kohler, 1986]. The canonical phonological structure of an utterance also affects rate perception. Koreman [2006] examined the rate perception of German speech samples that differed in their intended rates (i.e. related to the canonical structure of the phrase) and realized rates (i.e. referring to the actual realization of the phrase). Koreman [2006] showed that utterances with identical realized rates but different intended rates are perceived differently: the same realized rate with the faster intended rate is perceived as faster. Therefore, perceived speech rate is not only determined by the articulatory

events, but also by the listener's knowledge of the canonical structure of a particular utterance.

Other studies have examined whether speech rate perception depends on the language [e.g. French vs. English; Grosjean and Lass, 1977], on the linguistic task [e.g. task with or without comprehension activity; Grosjean, 1978], on the discourse style [read vs. spontaneous speech; Tjaden, 2000], and on the presence of language pathology [e.g. Tjaden, 2000]. Finally, it has been suggested that speech rate perception might vary as a function of extralinguistic factors, such as gender or relationship between speakers [e.g. Crown and Feldstein, 1991] or visual information [e.g. Greene, 1987].

Despite the centrality of the debate about the relationship between speech perception and production [see Diehl et al., 2004, for a review], the impact of the listeners' own rate of speech production on their perception of rate was given little attention. The conclusion drawn by Hoequist and Kohler [1986, p. 26], in their study of the perception of syllable duration, supports the suggestion of such a link: 'it [the discovery of tempo effects] should be seen as a confirmation of the more general idea that speech is a unified system, in which there can be no principled separation between speaker and hearer and between the systems they use for producing and receiving speech.' The purpose of the current research is thus to shed further light on the link between speech rate perception and production, specifically, to examine to what extent the listeners' own rate constitutes a determining factor in their perception of rate.

A few studies have attempted to tackle the question of the influence of rate production on rate perception. Lass and Cain [1972] investigated the hypothesis that English-speaking listeners' preferred speech rate depends on their actual speech rate: speakers who produced slow rates preferred listening to slow speech rates, and fast speakers tended to prefer fast rates. The authors concluded that the produced speech rate explained the variation in speech rate preference. This raises the question of whether speech rate production has an impact not only on speech rate preference, but also on speech rate perception. To our knowledge, only two studies have addressed this question and their conclusions are not wholly consistent.

Gósy [1992] made the assertion that the faster the speaker's own speech the less fast he perceives that of others as being so. She asked 37 Hungarian participants to classify as 'very slow', 'slow', 'moderate', 'accelerated', 'fast', and 'very fast' 12 synthesized sentences in Hungarian (the same sentence presented at different speech and articulation rates) and 9 natural speech samples in Hungarian (different monologues presenting various speech and articulation rates). The participants were divided into four groups according to their articulation and speech rates: slow ($n = 12$), moderate ($n = 9$), fast ($n = 10$) and rapid ($n = 6$) (this last group presented fast articulation rate but moderate speech rate). Gósy's [1992, p. 186] results showed 'a slight but definitive difference of tempo perception among the subjects with various speech tempo production'. More specifically, her main finding was that slow and fast speakers tended to judge speech rate in a similar way, while they presented different judgments in comparison with moderate and rapid speakers. She also reported that fast speakers, contrary to the other speakers, never classified any speech sample in the 'very fast' category, which supports her initial hypothesis that fast speakers tend to perceive fast speech as slower than the other speakers do. However, due to the unclear results and the lack of detailed statistical analysis, it is not easy to draw more precise conclusions about the differences in rate perception among the four groups of participants.

Koreman [2006] hypothesized, like Gósy [1992], that the listeners' own speaking habits may affect their perception of speech rate. In his experiment, 8 German participants with rate subjectively judged as fast, and 4 German participants with rate subjectively judged as slow, listened to 30 German phrases in which intended and realized rates differed (intended rates were computed on the basis of the canonical structure of the phrase, and realized rates on the basis of their actual realization). Participants heard different categories of speech (normal clear, fast clear, fast sloppy and slow clear), and they had to perform a scaling task, from -3 ('too slow') to 3 ('too fast'). Despite the fact that the differences between the speech categories in intended and realized phone rates were clearly perceived, results failed to show a difference between the rate perception in fast and slow listeners, and no interaction between the speech category and the participants' rate was found. This result tends to show that the listeners' own rate did not affect the perceived rate, and Koreman [2006] concluded that perceived speech rate does not seem to be a projection of the listeners' own speech behavior.

Methodological issues might explain the discordant results of Gósy [1992] and Koreman [2006]. Firstly, they opted for the categorization of the participants as a function of their judged speech rate (e.g. fast, moderate, slow speakers), rather than use quantitative measures of speech rate. Such an approach leads indeed to a loss of information about the individual differences, as well as to a loss of the effect size and the statistical power [see Brauer, 2002; MacCallum et al., 2002, for statistical arguments against categorization of quantitative data]. Secondly, sample sizes were relatively small, which might have decreased the statistical power of their analysis. Finally, Gósy's [1992] results may have been affected by the presence of large numbers of disfluencies in some of her speech samples.

The purpose of the present research is to examine the role of the listener's own rate in the perception of rate, taking into account these methodological issues. For this, we conducted a perception experiment, in which we used a magnitude estimation task, as this task has been shown to be appropriate for speech rate perception studies [see, for example, Lane and Grosjean, 1973; Cartwright and Lass, 1975]. French-speaking participants listened to and estimated readings in French that were recorded beforehand, at different speech rates (normal, fast and slow), by the same participants. The use of readings instead of spontaneous speech was motivated by the necessity of avoiding speech disfluencies, which could constitute a possible bias in the rate estimation task. The relation between the participant's own rate – which was not classified into a rate category, but considered as a quantitative measure – and the rate estimate was investigated by means of statistical analyses.

Method

Participants and Stimulus Materials

Twenty-eight French-speaking participants took part in this experiment. Their mean age was 27;9 years. The stimulus materials used in the perception experiment consisted of read versions of a text at normal, fast and slow rates, recorded by the 28 participants. We used the following updated French version of the 'Pop Fan Passage', which has been used in previous studies of 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.⁷

Recordings

Participants were recorded individually in a soundproofed booth. They were instructed to read the passage at normal, slow and fast speech rates. Recordings began with three readings at normal rate and continued with three readings at slow rate. After a short break, participants were asked to read the passage once again at normal rate and then 3 times at fast rate. Each participant's speech was recorded via microphone onto digital audio tape. The first reading at each rate and the normal reading after the break served respectively as training, and were not used in the perception experiment. Thus, there were two experimental readings at each rate by each speaker.

Measurements

For each rate (normal, slow and fast) the duration of speech intervals and pauses for both readings from each participant was measured with Praat 3.8 [Boersma, 2001]. Pauses were defined as silent intervals (sometimes breathing and other non-speech noises) longer than 200 ms. The beginning and the end of speech intervals were determined according to the following criteria: 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, sighs and non-speech noises were included in pauses. On the basis of these criteria, we measured the duration of all speech intervals and pauses, and we computed the total speech time (time from the beginning of the first speech interval to the end of the last one), the articulation time (sum of the duration of all speech intervals), the pause time (sum of the duration of all pauses), as well as the number of pauses.

Phonetic syllables (i.e. realized syllables) were counted in all productions on the basis of listening only, without resorting to spectrographic analyses. Following Grosjean and Deschamps [1975], we then computed speech rate (syll/min), articulation rate (syll/s)¹, as well as the number of pauses and the mean pause duration (ms) for the two readings at each rate (normal, slow and fast) for each participant.

Production Selection

We selected one of the two productions for each participant and for each rate according to the distribution of the temporal variables reported in Schwab [2007], where speech rate, articulation rate, pause number and duration, and speech rate range are examined for 40 participants (among them, the 28 participants of the present research) at three rates (normal, fast and slow). The selection of the 84 productions (28 speakers × 3 rates) to be used in the present perception experiment was done in such a way that the distribution of the selected normal, fast and slow productions, as well as the distribution of speech rate range (difference between slow and fast speech rates), matched as close as possible the distributions of the 40 productions in Schwab [2007; see Schwab, 2010 for details]. Table 1 shows the mean value and standard deviation (in parenthesis) of temporal variables (speech rate, articulation rate, number of pauses and mean pause duration) at slow, normal and fast rates, and speech rate range.

Figure 1 presents the distribution of speech rates (slow, normal and fast) and the distribution of rate range. There is a significant rate difference [$F(2, 54) = 243.83, p < 0.0001$] between slow, normal and fast intended rates, and there are significant differences between each rate (Tukey HSD, $p < 0.01$ for all comparisons). Note that the three speech rates overlap (fig. 1): slow rate of some participants corresponds to a normal or even a fast rate for other speakers. Conversely, the fast rate of some speakers is closer to the normal rate of other participants.

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 and Deschamps, 1975]. Speech rate can be expressed in syllables per minute [Grosjean and Deschamps, 1975], in words per minute [Goldman-Eisler, 1968] or in phones per second [Gósy, 1992], while articulation rate is generally expressed in syllables per second [Grosjean and Deschamps, 1975] or in phones per second [Koreman, 2006].

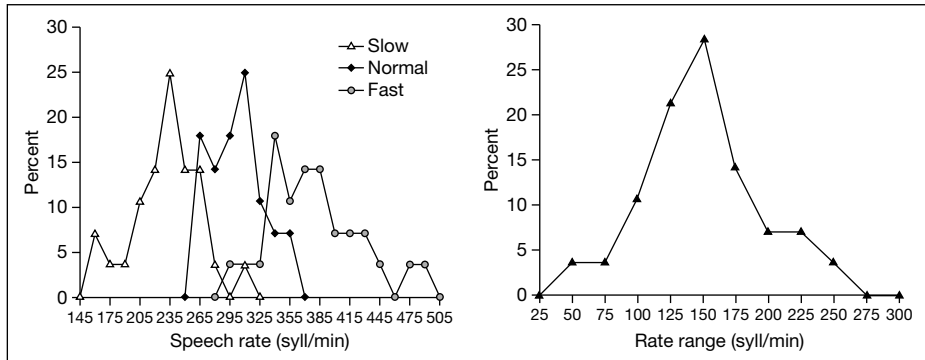


Fig. 1. Distribution of speech rates (slow, normal and fast) and speech rate range distribution (n = 28).

Table 1. Mean value and standard deviation (in parentheses) of temporal variables at slow, normal and fast rates, and speech rate range

Temporal variables	Speech rate		
	slow	normal	fast
Speech rate, syll/min	232 (34)	301 (26)	379 (46)
Articulation rate, syll/s	4.7 (0.5)	5.5 (0.4)	6.6 (0.6)
Number of pauses	6.9 (2.5)	3.5 (1.5)	1.6 (1.4)
Mean pause duration, ms	522 (111)	413 (76)	345 (38)
	Speech rate range, syll/min		
	147 (43)		

Procedure

The 84 productions (28 speakers × 3 rates) were split up into three lists (A, B and C), in such a way that the three productions of each speaker appeared in a different list. Within each list, the productions were ordered so that no more than two with the same rates (normal, slow or fast) followed each other. Moreover, we added one filler production at the beginning of each list, and three filler productions were chosen as practice.

Participants were run individually or in pairs. After listening to each production through headphones, they were instructed to perform a magnitude estimation task [see Stevens, 1957, 1975, for details]. More specifically, the participant had to assign a number to each speech rate he heard. The number 10 corresponded to what the participant considered a normal speech rate. The number he gave had to be proportional to what he thought was a normal speech rate (10). For example, 20 corresponded to a speech rate which was twice as fast as what the participant considered a normal speech rate, and 5 corresponded to a speech rate which was half the speed of the normal speech rate. No explicit instructions were given about what should be considered as a normal rate, and no reference rate was presented as a basis for estimation.

Each session consisted of the presentation of the three practice productions, followed by the presentation of the three lists (84 productions), with a break between each list. Half the participants (group 1) heard the three lists in the order A, B and C, and the other half (group 2) heard them in the reverse order (C, B and A). The experiment lasted approximately 75 min.

We collected the speech rate estimate given by the 28 participants for the 28 productions heard for each rate (2,352 data points in total, 784 data points per rate). We first examined the possible presence of a list order effect on rate estimates. We found no significant interaction between participant group (1 and 2) and list (A, B, C) [$F(2, 2,346) = 0.04$, n.s.], which indicates that the list order had no effect on rate estimation.

Then, we computed, within each rate, the mean estimate for each participant (28 data points per rate). We used logarithmic values in all statistical analyses, given that, as mentioned in the 'Introduction', speech rate perception follows Stevens' power function law [Cartwright and Lass, 1975], but for the sake of clarity we present raw data means in the following sections.

Results and Discussion

The key question was whether the listener's own rate affects his estimation of other speakers' rates. To answer this question, we first studied the relationship between the listener's own rate and his mean rate estimate. Second, we examined the moderator effect of the listener's own rate on his rate estimate, in order to determine to what extent the listener's own rate affects his perception of rate differences.

Relationship between the Listener's Own Rate and His Mean Rate Estimate

Figure 2 represents the rate estimate as a function of the listener's own rate for normal, fast and slow rates. Within each rate, each data point represents the listener's mean estimate (across the 28 samples he heard) as a function of his own rate. Listeners were able to distinguish between the three rates: they gave estimates that differ significantly between the rates [normal = 10.82, fast = 17.91, slow = 6.07; $F(2, 54) = 488.84$, $p < 0.001$; Tukey HSD, $p < 0.01$ for all comparisons].

Secondly, and more interestingly, as can be seen in figure 2, we found a negative correlation between mean rate estimate and own rate at normal and slow rates (respectively, $r = -0.45$, $p < 0.05$; $r = -0.39$, $p < 0.05$), but no correlation at fast rate ($r = -0.11$, n.s.). These results show that speakers with a slow speech rate tended to overestimate the sample speech rates (i.e. to give a higher estimate) in comparison to fast speakers, at normal and slow rates.

The lack of a relationship at fast rates might come from the speakers' higher variability in how much they increase their rate in comparison with how much they decrease it. If so, the correlation between the speakers' own fast and normal rates should be weaker than the correlation between the speakers' own slow and normal rates. Nevertheless, we found very similar correlations (respectively, $r = 0.67$, $p < 0.001$; $r = 0.66$, $p < 0.001$; $z = -0.007$, n.s.). Furthermore, no correlation emerged between the listeners' normal rate and their estimate of fast rates ($r = 0.25$, n.s.). Thus, the lack of a relationship at a fast speech rate is not related to speakers' ability in how much they increase and decrease their rate.

Moreover, we examined whether speech rate estimation was rather related to the listeners' own articulation rate (i.e. the rate at which they articulate utterances, without taking into account the pauses they produce). No significant correlation was found between the listeners' articulation rate and their speech rate estimate, whatever the rate

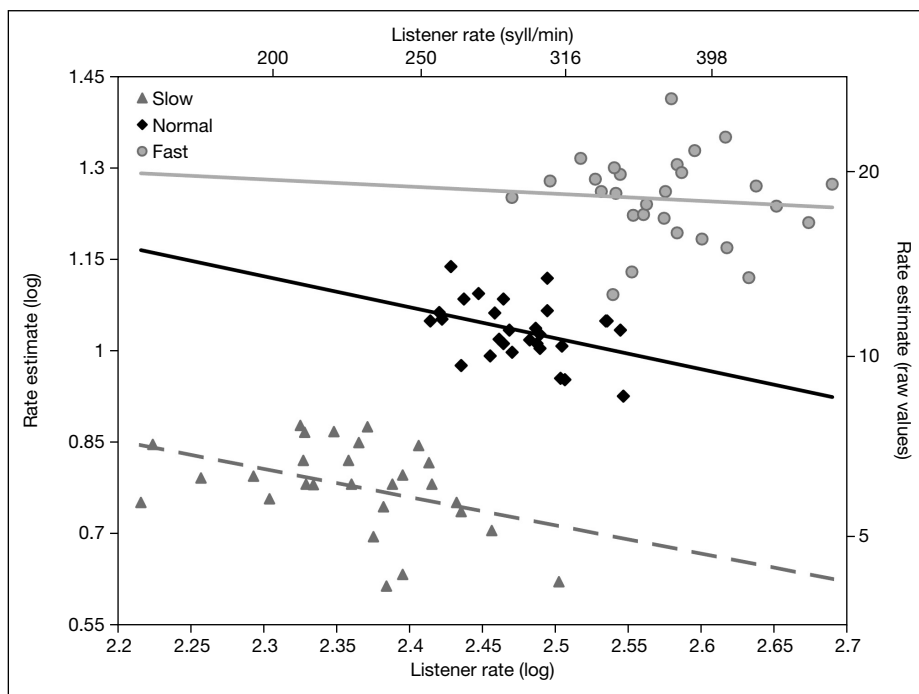


Fig. 2. Rate estimate as a function of the listener's own rate for normal, fast and slow rates.

(normal, fast and slow) may be, although the correlation for normal rate just failed to reach significance (normal: $r = -0.36$, $p = 0.06$; fast: $r = -0.06$, n.s.; slow: $r = -0.27$, n.s.). Nevertheless, it should be noted that these correlations are not significantly different from the aforementioned correlations between the listeners' speech rate and speech rate estimate (normal: $z = -0.45$, n.s.; fast: $z = -0.21$, n.s.; slow: $z = -0.56$, n.s.).

Our results confirm the hypothesis of a relationship between speech rate production and speech rate perception. However, this relation exists at normal and slow rates but not at fast rate. Further investigation is needed to examine in more detail the reasons leading to the absence of a correlation between speech rate production and speech rate perception for fast rates. In particular the effect of fast speech cues, such as reductions [Trouvain, 1999] and sloppy speech [Koreman, 2006], should be explored in order to understand why fast speech is perceived as fast whatever the listener's own rate may be.

Moderator Effect of the Listener's Own Rate in His Rate Estimate

Further analyses were conducted in order to determine how the relationship between actual rate and rate estimate varied as a function of the listener's speech rate. In other words, we examined the influence of the listener's own rate on his sensitivity to rate differences. Thus we studied to what extent the listener's rate is a moderator in the estimation of rate. According to Baron and Kenny [1986, p. 1174], 'a moderator is a qualitative [. . .] or quantitative [. . .] variable that affects the direction and/or strength

of the relation between an independent or predictor variable and a dependent or criterion variable’.

Considering the moderator hypothesis, Rate estimate (the dependent variable) may be affected by Actual rate as a predictor, by Listener rate as a moderator, and by the interaction of these two (Actual rate \times Listener rate, a variable called ‘Interaction’ hereafter). Baron and Kenny [1986, p. 1174] argue that the moderator hypothesis would be supported if the interaction is significant, and they specify that the significant main effects for the predictor (i.e. Actual rate) and the moderator (i.e. Listener rate) ‘are not directly relevant conceptually to testing the moderator hypothesis’.

We ran three separate regression analyses for the normal, fast and slow rates, with Rate estimate as a dependent variable, and Actual rate, Listener rate and Interaction as predictors. In view of the fact that the predictors (Actual rate, Listener rate and Interaction) were highly correlated (Variance Inflation Factor, VIF >10)², Listener rate was removed from the analyses, given that its main effect was not relevant for the moderator hypothesis we wanted to test. On the other hand, Actual rate was included in the analyses as we were interested in examining how the listeners perceived the changes in the actual rates. The regression analyses were performed with Rate estimate as the dependent variable and Actual rate and Interaction as predictors (VIF = 2), separately for the normal, fast and slow rates³.

As expected, the regression analyses showed an effect of Actual rate, at all rates [normal: $\beta = 3.04$, $t(781) = 21.05$, $p < 0.001$; fast: $\beta = 1.88$, $t(781) = 19.25$, $p < 0.001$; slow: $\beta = 2.09$, $t(781) = 24.01$, $p < 0.001$], meaning that Actual rate had a strong impact on Rate estimate. In other words, listeners were able to perceive rate differences within normal, fast and slow speech samples, respectively. This finding, taken together with the results presented in the previous section, supports the idea that listeners are able to distinguish not only the three categories of rate (normal, fast and slow), but also the different rates within each category.

More interestingly, results showed an effect of Interaction, at all rates [normal: $\beta = -0.23$, $t(781) = -5.67$, $p < 0.001$; fast: $\beta = -0.06$, $t(781) = -2.06$, $p < 0.05$; slow: $\beta = -0.23$, $t(781) = -6.46$, $p < 0.001$], which supports the hypothesis of the moderator effect of the listener’s own rate in his rate estimate, although with a larger effect at normal and slow rates than at fast rate. The faster a listener speaks, the less his rate estimates rise as a function of actual rates. In other words, the faster the listener’s own rate, the weaker the effect of the actual rates on rate estimate.

A graphic illustration of the conditional relationship between Rate estimate and Actual rate on Listener rate (shown by Interaction) is given in figure 3. In fact, it shows how the relationship between Rate estimate and Actual rate (i.e. conditional slope) varies conditional on Listener rate, or, in other words, how the listeners perceive the rate differences according to their own rate. As can be seen, for all rates, the conditional slope decreases as a function of Listener rate, meaning that listeners with a fast own rate are less sensitive to rate differences than listeners with a slower rate. We can also

² The 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 et al., 2003, p. 423].

³ The regression equation presents the following shape: $\hat{y} = \beta_1x + \beta_2xz + \beta_0$, where x represents the Actual rate, z the Listener rate; thus, xz represents the interaction between Actual rate and Listener rate. Algebraic transformation leads to the following equation $\hat{y} = (\beta_1 + \beta_2z)x + \beta_0$, in which the regression coefficient $(\beta_1 + \beta_2z)$ corresponds to the conditional slope for Rate estimate regressed on Actual rate conditional on Listener rate.

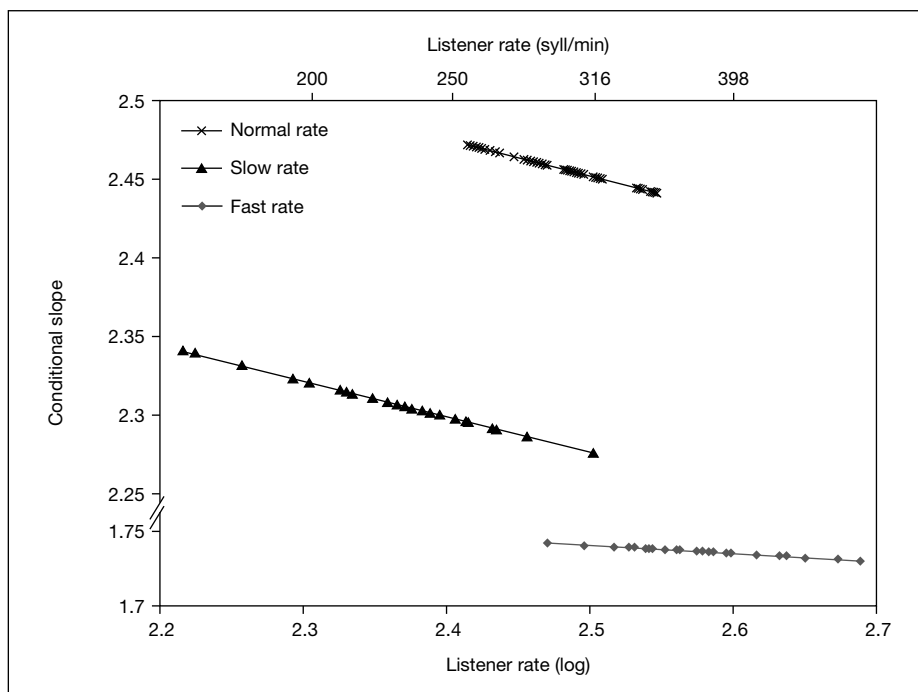


Fig. 3. Conditional slope of Rate estimate on Actual rate (i.e. relationship between Rate estimate and Actual rate conditional on Listener rate) as a function of Listener rate, for normal, fast and slow rates.

observe that the slope falls more steeply at normal and slow rates than at fast rates, showing again that Listener rate has a smaller effect on rate estimate at fast rates.⁴

We also investigated, with the same statistical procedure, whether speech rate estimation could be explained by the actual articulation rate and by the listener's articulation rate. As for speech rate, we found an effect of actual articulation rate on speech rate estimate at all rates [normal: $\beta = 2.88$, $t(781) = 15.62$, $p < 0.001$; fast: $\beta = 2.20$, $t(781) = 17.87$, $p < 0.001$; slow: $\beta = 2.96$, $t(781) = 16.34$, $p < 0.001$]. Moreover, results showed an interaction between the actual articulation rate and the listener's articulation rate at normal and slow rates [normal: $\beta = -0.70$, $t(781) = -4.00$, $p < 0.001$; slow: $\beta = -0.74$, $t(781) = -3.92$, $p < 0.001$], but not at fast rate [$\beta = -0.12$, $t(781) = -1.11$, n.s.]. These results are consistent with the ones previously reported on speech rate: the faster the listener's own articulation rate, the weaker the effect of the actual articulation rate on rate estimate, at normal and slow rates. Regarding fast rate, while the listener's speech rate played a small role (although significant) in rate estimation, the listener's articulation rate plays no significant role.

Finally, we conducted an analysis to rule out the possible effect of listeners judging their own voice (recall that each listener was among the 28 speakers), given that some of the listeners might have recognized their own voice, while others might not. Therefore

⁴ We also ran the same analyses for slow and fast rates with the listener's normal rate as moderator. Results showed an effect of Interaction, at both rates [slow: $\beta = -0.47$, $t(781) = -7.57$, $p < 0.001$; fast: $\beta = -0.18$, $t(781) = -4.92$, $p < 0.001$], with again a smaller effect of the listener's own rate at fast rate.

we excluded from the dataset own voice judgments, and applied the same statistical methodology as previously described. Results showed very similar trends as with the inclusion of own voice judgments: the interaction between the actual speech rate and the listener's own rate was present at normal and slow rates [normal: $\beta = -0.23$, $t(753) = -5.46$, $p < 0.001$; slow: $\beta = -0.23$, $t(753) = -6.41$, $p < 0.001$] and marginal at fast rate [$\beta = -0.05$, $t(753) = -1.9$, $p = 0.06$]. Thus, the fact that the listeners judged their own voice did not seem to affect the influence their own rate had on their rate estimation.

General Discussion

The purpose of this investigation was to examine the role that speech rate production might play in speech rate perception. Participants were first asked to read a passage at normal, fast and slow rates. They were then instructed to estimate the various speech samples they all produced, using a magnitude estimation task.

Listeners were able to distinguish between normal, fast and slow produced rates, despite the slight overlap between rates in the three intended rate categories. One explanation might be that they base their estimation of rate not only on temporal aspects, but also on the different cues which are closely linked to speech rate variation. Change in speech rate produces a complex, nonlinear expansion or compression of the speech signal, affecting many speech acoustic-phonetic properties [e.g. Zellner, 1998]. At the segmental level, speech rate variation affects the spectral and temporal properties of the segments [e.g. Volaitis and Miller, 1992; Moon and Lindblom, 1994; Allen and Miller, 1999]. It might also involve a phonological reorganization of the utterances, which results in addition (e.g. schwa epenthesis at slow rate), suppression (e.g. schwa deletion at fast rate) or different realization (e.g. resyllabation) of syllables [e.g. Zellner, 1998]. And, at the suprasegmental level, speech rate has been shown to have an impact on the phonetic F0 realization (e.g. modification of melodic contours, pitch range and displacement, F0 resets), and on the prosodic structure of the utterance (e.g. changes in the strength of prosodic boundaries) [e.g. Fougeron and Jun, 1998; Trouvain and Grice, 1999]. Therefore, besides temporal information, the segmental and suprasegmental consequences of speech variation might play a role in the ability of the listeners to distinguish, for example, slow and normal rates that overlapped.

Secondly, results showed that the listener's own rate has two effects on rate perception. First, listeners with a slow speech rate tend to overestimate the sample rates (i.e. give a higher estimate) in comparison to fast speakers (at normal and slow rates, at least). Second, listeners with a slow speech rate are more sensitive to rate differences within a given rate category (normal, slow and fast) than listeners with a fast rate.

It should be noted that at fast rates, the effect of the listener's own rate is weaker than at normal and slow rates. It is likely that the aforementioned segmental and prosodic cues affected by rate variation are more salient at fast rate, and that they play a more important role than the temporal cues in the perception of fast rate. Further investigation, using, for example, Koreman's [2006] methodology (intended vs. realized speech), needs to be conducted in order to confirm this hypothesis.

Moreover, our results showed that the perception of speech rate is affected by the listener's own rate whether we consider his speech rate or his articulation rate. This finding supports the conclusion of Grosjean and Lane [1976], who showed that articulation rate plays a determining role in the perception of speech rate.

An interesting question that may arise from our results concerns the direction of the relationship between rate production and perception. We hypothesized, like other researchers [Gósy, 1992; Koreman, 2006], that rate production influences rate perception, but we did not consider the reverse interpretation. However, at the segmental level, the direction of the link between production and perception has been considered both ways. Following the hypothesis of Perkell et al. [2004], speech perception affects speech production, while according to other researchers [Paliwal et al., 1983], speech production regulates speech perception. At the suprasegmental level, further studies are clearly needed to determine the possible influence of rate perception on rate production.

Finally, it would be of great interest to examine whether speech rate perception is affected not only by the listener's actual rate, but also by the listener's perception of his own rate (i.e. the perception of autophonic rate). We have shown that the fact that the listeners judged their own productions had no impact on the influence that their own rates have on their rate estimates. But we cannot know whether the listeners recognized themselves, and whether they knew that they were judging their own productions. Indeed, it is well known that people hearing their own voice often fail to recognize themselves. For example, Rousey and Holzman [1967] showed that only 38% of the subjects recognized their own voice. Thus, given the fact that there is no way in our experiment to make sure that listeners estimated their own rate in a conscious way – knowing that they were listening to themselves –, the interesting issue about the role of the perception of autophonic rate in the perception of extraphonic rate should be the subject of further studies.

As mentioned in the 'Introduction', the listener's perceptual impression of speech rate seems to be influenced not only by the temporal properties of speech, but also by other linguistic and extralinguistic factors. The present research confirms this assumption, and adds another factor – the listener's own rate – to the list of the non-temporal variables capable of affecting speech rate perception.

Acknowledgments

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. Thanks also to Audrey Buerki Foschini and Laurence White for their valuable comments on an earlier version of this article.

References

- Allen, J.S.; Miller, J.L.: Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *J. acoust. Soc. Am.* 106: 2031–2039 (1999).
- Baron, R.M.; Kenny, D.A.: The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Personality social Psychol.* 51: 1173–1182 (1986).
- Boersma, P.: Praat, a system for doing phonetics by computer. *Glott int.* 5: 341–345 (2001).
- Brauer, M.: L'analyse des variables indépendantes continues et catégorielles: alternatives à la dichotomisation. *Année psychol.* 102: 449–484 (2002).
- Cartwright, L.R.; Lass, N.J.: A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling. *Lang. Speech* 18: 358–365 (1975).
- Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S.: *Applied multiple regression/correlation analysis for the behavioral sciences*; 3rd ed. (Erlbaum, Mahwah 2003).
- Crown, C.L.; Feldstein, S.: The perception of speech rate from the sound-silence patterns of monologues. *J. psycholinguist. Res.* 20: 47–63 (1991).

- den Os, E.: Perception of speech rate of Dutch and Italian utterances. *Phonetica* 4: 124–134 (1985).
- Diehl, R.L.; Lotto, A.J.; Holt, L.L.: Speech perception. *Annu. Rev. Psychol.* 55: 149–179 (2004).
- Feldstein, S.; Bond, R.: Perception of speech rate as a function of vocal intensity and frequency. *Lang. Speech* 24: 387–395 (1981).
- Fougeron, C.; Jun, S.-A.: Rate effects on French intonation: prosodic organization and phonetic realization. *J. Phonet.* 26: 45–69 (1998).
- Goldman-Eisler, F.: *Psycholinguistics: experiments in spontaneous speech* (Academic Press, London 1968).
- Gósy, M.: The perception of suprasegmentals; in Gósy, *Speech perception*, pp. 159–218 (Hector, Frankfurt am Main 1992).
- Green, K.P.: The perception of speaking rate using visual information from a talker's face. *Perception Psychophysics* 42: 587–593 (1987).
- Grosjean, F.: 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 (1972).
- Grosjean, F.: The perception of rate in spoken and sign language. *Perception Psychophysics* 22: 408–413 (1977).
- Grosjean, F.: Perception of rate and processing of sentences (Northeastern University, Boston, unpublished, 1978).
- Grosjean, F.; Deschamps, A.: Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31: 144–184 (1975).
- Grosjean, F.; Lane, H.: How the listener integrates the components of speaking rate. *J. exp. Psychol.* 2: 538–543 (1976).
- Grosjean, F.; Lass, N.: Some factors affecting the perception of reading rate in English and in French. *Lang. Speech* 20: 198–208 (1977).
- Hoequist, C.E.; Kohler, K.J.: Summary of speech rate perception research at Kiel. *AIPUK* 22: 5–29 (1986).
- Kohler, K.J.: Parameters of speech rate perception in German words and sentences: duration, F0 movement, and F0 level. *Lang. Speech* 49: 115–139 (1986).
- Koreman, J.: Perceived speech rate: the effect of articulation rate and speaking style in spontaneous speech. *J. acoust. Soc. Am.* 119: 582–596 (2006).
- Lane, H.; Grosjean, F.: Perception of reading rate by speakers and listeners. *J. exp. Psychol.* 97: 141–147 (1973).
- Lass, N.J.; Cain, C.J.: A correlational study of listening rate preferences and listeners' oral reading rates. *J. audit. Res.* 12: 308–312 (1972).
- MacCallum, R.X.; Zhang, S.; Preacher, K.J.; Rucker, D.D.: On the practice of dichotomization of quantitative variables. *Psychol. Meth.* 7: 1–40 (2002).
- Moon, S.-J.; Lindblom, B.: Interaction between duration, context and speaking-style in English stressed vowels. *J. acoust. Soc. Am.* 96: 40–55 (1994).
- Paliwal, K.K.; Lindsay, D.; Ainsworth, W.A.: Correlation between production and perception of English vowels. *J. Phonet.* 11: 77–83 (1983).
- Perkell, J.S.; Guenther, F.H.; Lane, H.; Matthies, M.L.; Stockmann, E.; Tiede, M.; Zandipour, M.: The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *J. acoust. Soc. Am.* 116: 2338–2344 (2004).
- Pfzinger, H.R.: Local speech rate perception in German speech. *Proc ICPhS, San Francisco 1999*, pp. 893–896.
- Putman, W.B.; Street, R.L.: The conception and perception of noncontent speech performance: implications for speech accommodation theory. *Int. J. Sociology Lang.* 46: 97 (1984).
- Quené, H.: On the just noticeable difference for tempo in speech. *J. Phonet.* 35: 353–362 (2007).
- Rietveld, A.; Gussenhoven, C.: Perceived speech rate and intonation. *J. Phonet.* 15: 273–285 (1987).
- Rousey, C.; Holzman, P.S.: Recognition of one's own voice. *J. Personality soc. Psychol.* 6: 464–466 (1967).
- Schirmer, A.: Timing speech: a review of lesion and neuroimaging findings. *Cognitive Brain Res.* 21: 269–287 (2004).
- Schwab, S.: Les variables temporelles dans la production et la perception de la parole; thèse de doctorat Université de Genève (2007).
- Schwab, S.: Does a talker's own rate of speech affect his/her perception of others' speech rate? In Schmid, Schwarzenbach, Studer, *La dimensione temporale del parlato. Atti del 5° Convegno Nazionale AISV, Università di Zurigo, 2009*, pp. 445–454 (EDK Editore, Torriana 2010).
- Stevens, S.S.: On the psychophysical law. *Psychol. Rev.* 64: 153–181 (1957).
- Stevens, S.S.: *Psychophysics: introduction to its perceptual, neural, and social prospects* (Wiley, New York 1975).
- Street, R.L.; Brady, R.M.: Speech rate acceptance ranges as a function of evaluation domain, listener speech rate, and communication context. *Commun. Monogr.* 49: 290–308 (1982).
- Tjaden, C.: A preliminary study of factors influencing perception of articulatory rate in Parkinson disease. *J. Speech Lang. Hear. Res.* 43: 997–1010 (2000).
- Trouvain, J.: Phonological aspects of reading rate strategies. *Phonus* 4: 15–35 (1999).
- Trouvain, J.; Grice, M.: The effect of tempo on prosodic structure. *Proc ICPhS, San Francisco 1999*, pp. 1067–1070.
- Volaitis, L.E.; Miller, J.L.: Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of phonetic categories. *J. acoust. Soc. Am.* 92: 723–735 (1992).
- Zellner, B.: Caractérisation et prédiction du débit de parole en français: une étude de cas; thèse de doctorat Université de Lausanne (1998).