

# What is the Just Noticeable Difference for tempo in speech?

Hugo Quené

Utrecht University

## Abstract

Tempo (speaking rate) varies both between and within speakers. Such variations in tempo are easily noticeable. But what is the just noticeable difference for tempo in speech? As a first approximation, between-speaker tempo variation is quantified in a sample of similar interviews with 80 speakers of Dutch. Second, the JND is assessed using a somewhat unconventional method, viz. detection of tempo drift. This results in a JND of 15%; it is argued that this value is upwardly biased. Third, the JND is assessed using a conventional same~different pairwise comparison, yielding a JND of about 10%. Tempo variations between and within speakers typically exceed this JND, and are therefore potentially important in speech communication.

## 1 Introduction

Human speech is produced by moving the vocal organs and articulators. These movements result in an articulated speech signal, in which phonetic events occur at particular moments in time. The rate at which these speech events occur constitutes the tempo or speed or rate of speech. Many textbooks in phonetics state that speakers vary their speaking rate, in anticipation of the time listeners will need to process their words. Hence, important or unpredictable portions are spoken at a relatively slower rate (e.g., Zwaardemaker & Eijkman, 1928, p.304; Nootboom & Cohen, 1984, p.165). This tendency follows from the adaptation principle or H&H principle (Lindblom, 1989): speakers adjust phonetic properties of their speech to ensure an optimal balance between economy of articulatory energy, and perceptual clarity for the listener. After all, speakers speak in order to be understood. This phonetic principle underlies the usual rhetoric advice to public speakers, to slow down when important information is conveyed (e.g., Humes, 2002). In most phonetic studies supporting these claims and recommendations, however, information value (newness, importance) and accentuation have been confounded. Eefting and Nootboom (1993) (and Nootboom & Eefting, 1994) show that speaking rate is *not* slower for new words, if accentuation is also taken into account (at least for the single professional standard speaker in their study). Their conclusion is based on the relatively small increment by about 4% in duration of a target word *John*, in the [+new, –accent] condition (1) relative to the [–new, –accent] condition (2).

(1) (What did you say?) John MILLER is ill.

(2) (What did you say about John Miller?) John Miller is ILL.

Hence, this evidence rests on the implicit assumption that the observed increment of 4% in duration (or decrement in speaking rate) is not relevant for speech communication<sup>1</sup>. If this 4% difference is above the difference limen for speaking rate, however, then this assumption

---

<sup>1</sup> By way of comparison: accenting the target word ('JOHN is ill') yields an increment in word duration and in syllable duration of about 25% (Eefting & Nootboom, 1993).

is not warranted. But what exactly is the difference limen (DL) or just noticeable difference (JND) for speaking rate? If a speaker changes tempo, how large does the tempo change have to be in order to be perceptually relevant? To this date, a few studies have addressed this question (Benguerel & D'Arcy, 1986; Eefting & Rietveld, 1989; Nooteboom & Eefting, 1994), but these studies leave considerable room for improvement, as will be discussed below.

As a first approximation to an answer, we could inspect tempo differences between speakers. If speakers are observed to differ in their basic tempo (e.g., Goldman-Eisler, 1968; Den Os, 1985; Van Heuven, 2003), then these between-speaker differences must exceed the observer's JND. Hence, the range of tempi in a large speech corpus can help us to establish the JND for speech tempo. The variation in speaking rate was investigated by means of the Corpus of Spoken Dutch for this purpose.

## 2 Corpus analysis

The Corpus of Spoken Dutch (CGN, e.g. Oostdijk, 2000) was used to quantify the between-speaker variation in speaking rate. For this purpose, we concentrated on the sub-corpus containing interviews with high-school teachers of Dutch in the Netherlands and Belgium (Van Hout et al., 1999). Only the speakers from the Netherlands are discussed here. The relevant sub-corpus contains interviews with 80 speakers, each interview lasting about 15 minutes. Interviewed speakers ('interviewees') were stratified by dialect region (four regions within the Netherlands), sex, and age group (below 35 vs. over 45 years of age), with  $n=5$  speakers in each cell. All 80 speakers are assumed to speak a variety of Standard Dutch as used in the Netherlands. All 80 interviews were conducted by the same interviewer (female, age 26), and similar topics were discussed across interviews. Hence, language variety, conversation partner, and conversation topic were eliminated as confounding factors.

For each interview, the orthographic transcript of the interviewee was extracted, and the speaking time of the interviewee was determined from the time marks in this transcript. Pauses were thus excluded from the interviewee's speaking time. On average, the interviewee spoke during about 3/4 of the total interview duration. Phonetic speaking rate is usually expressed in a syllables-per-second scale (e.g., Stetson, 1988) or average syllable duration (Goldman-Eisler, 1968). Because the necessary phonetic transcripts are not available for this part of the CGN, the words-per-minute (wpm) scale was used instead. The number of words spoken by the interviewee was counted in the orthographic transcripts, using standard word-processing software (TextPad 4.7.2). For each interview, the speaking rate was calculated from the speaking time and word count.

In these 80 interviews, the average speaking rates of the interviewees range between 151 and 281 words per minute, with an overall average of 220 wpm ( $s=25$ ; the distribution is approximately normal,  $KS=0.068$ ,  $p=.45$ ). This range and variation between speakers is quite large, considering that the interviews were similar with respect to interviewer, topics discussed, and total duration of interview. An average speaker would require 4.55 minutes to produce 1000 words. The fastest speaker would require only 3.56 minutes ( $0.78 \times$  average), and the slowest speaker 6.62 minutes ( $1.45 \times$  average).

Further analysis of variance shows that the speaking rate varies significantly between male and female speakers [227 vs. 213 wpm, respectively;  $F(1,76)=15.0$ ,  $p<.001$ ], and between younger speakers (mean age 33.5 year; mean rate 211 wpm) and older speakers [mean age 51.6 years; mean rate 230 wpm;  $F(1,76)=7.7$ ,  $p=.007$ ]. No interaction was observed between the sex and age-group factors,  $F(1,76)=1.2$ , n.s.

Hence, speaking rates do vary considerably between speakers, as has been observed many times before (e.g. Goldman-Eisler, 1968; Den Os, 1985). The rate of the fastest speaker is almost double that of the slowest speaker, in this part of the CGN. It comes as no surprise, then, that these large between-speaker differences are highly noticeable. A rough estimate of JND is obtained from the normal distribution of speaking rates across speakers, as follows. Let us make the debatable assumption that perception mirrors production, and that listeners' JND corresponds to 1 standard deviation of the distribution of produced tempi. This would amount to an estimated JND of 25/220 or 11%. In reality, however, listeners seem to perform better in speech tempo discrimination (Nooteboom & Eefting, 1994), and hence the JND is smaller, by some unknown amount.

### 3 On the Just Noticeable Difference for tempo

In the preceding section, it was assumed that listeners discriminate between the slowest and fastest decatiles of the population. This assumption was necessary because the true JND for speech tempo remains to be determined. Hence the main question of this study: what is the just noticeable difference (JND) for tempo in speech? How much do speaking rates have to deviate from a reference in order to become noticeable, and hence relevant in speech communication? In the following sections, two experiments are reported that were aimed at establishing the JND for tempo in speech.

Research on JNDs for tempo has concentrated on *music* perception. Ellis (1991) presented listeners with a 6-bar, 24-beat musical fragment at various base tempi. After a stable period (of random duration), the tempo of the fragment started to drift gradually (up or down, with +2% or -2% on each subsequent beat, to extremes of either +16% or -10%). Using the staircase adjustment method, JNDs of 5.1% to 13.9% change in tempo were found, depending on the direction of drift and on the base tempo<sup>2</sup>.

Drake and Botte (1993, Experiment 3) presented listeners with two 5-tone sequences to compare (2IFC paradigm). Using the staircase adjustment method, they found JNDs for this type of stimuli to be 6% to 10% for nonmusical listeners, and 3% to 8% for musicians, with the lowest JND at a base rate of 100 beats per minute (inter-onset interval 0.6 s). In an ERP study having a similar design, Pfeuty, Ragot and Pouthas (2003) report that a 4% change in inter-onset interval in a 7-tone sequence yields a discriminability value  $d'$  of 1.52, which suggests that the JND is smaller than 4% in their experiment.

Levitin and Cook (1996) cite an unpublished study by Perron (1994) involving computer sequencers or drum machines, as used in popular music. Although such machines turned out to have an average tempo deviation of 3.5%, most listeners do not notice these deviations (and neither do professional drummers). This suggests that the JND for musical tempo is at least 3.5%.

These studies (and others not discussed here) indicate that the JND for musical tempo is approximately 6% to 8% of the base tempo. For *speech* tempo, however, only a few estimates of JND are available. Benguerel and D'Arcy (1986, Experiment 4) presented a few pre-selected listeners with reiterant speech stimuli [nananananana], with exponentially decreasing or increasing duration of each subsequent syllable. Listeners judged sequences as "regular", even for decelerating sequences with increasing syllable duration. However, a conventional

---

<sup>2</sup> This could be considered as a 2IFC paradigm, in which the beginning and the ending of the musical fragment constitute the two intervals to be compared.

JND value cannot be derived easily from their results. In addition, it is not clear whether and how this generalizes to normal speech and to other listeners.

Eefting and Rietveld (1989) presented listeners with two versions of a speech utterance (2IFC paradigm); tempo was always unchanged in one version. The reported JND of 4.4% is remarkably low, and even lower than some values reported above for musical tempo. Eefting and Rietveld argue that this may be due to listeners' adaptation to the stimuli. Interestingly, a JND this low would render the 4% change in tempo reported by Eefting and Nootboom (1993) perceptually relevant. Again, generalization to other speech stimuli is questionable.

Assessing JNDs for tempo is surely more difficult in speech than in music, where presenting several intervals between beats requires only a few seconds. In speech, each syllable has its intrinsic duration and temporal structure, depending on the articulatory gestures involved in producing that syllable. This obscures the temporal regularity to some extent. Consequently, listeners may well need more than a few (stressed) syllables to determine the speaking rate in a speech fragment.

The classical method to determine a JND is to present two stimuli shortly after each other, which are either identical or slightly different in one property, and to ask listeners whether these are the same or different (2IAX paradigm) (e.g. Nootboom & Eefting, 1994), or which one of the two has more of the property under investigation (2IFC paradigm) (e.g. Den Os, 1985; Drake & Botte, 1993). Hence the assumption is that listeners have an active memory trace of *both* stimuli when they respond. This assumption may well be unwarranted if longer stimuli are used, such as musical melodies or speech fragments that typically last at least a few seconds. An experimental paradigm involving a single stimulus presentation is to be preferred in these cases.

Hence, it will be attempted below to establish the JND for speaking tempo, by means of drift detection. In this task, the tempo in a stimulus starts to drift (i.e., to accelerate or decelerate), and the listeners' task is to respond as fast as possible to a perceived change in tempo. The JND is then calculated from the amount of change in the stimulus, at the time the listener responds. This method has already been successful for investigating musical tempo (Ellis, 1991; Grondin, 2003).

## 4 Detection of tempo drift

### 4.1 Method

Stimulus materials consisted of 36 text passages that resembled short news items, consumer reports, personal anecdotes, etc. Four female speakers spoke nine passages each, at a normal rate. Their productions were recorded on DAT, and re-digitized (16 kHz, 16 bits). Each passage contained about 10 s of speech in two or three sentences.

Between recorded text passages, the onset of tempo drift was varied between 2 and 4 s after onset of speech, to avoid strategic effects. PSOLA manipulation was used to create both the accelerating and the decelerating version of each recording. Relative duration of the speech recording was compressed (by linear interpolation from 1.0× to an extreme of 0.8× original; tempo acceleration) and expanded (to 1.2×; tempo deceleration) over a 5-s time window<sup>3</sup>. The two versions were counterbalanced over two stimulus lists, each of which also contained four practice items. Within each stimulus list, the blocking order of accelerating and decelerating items was also counterbalanced.

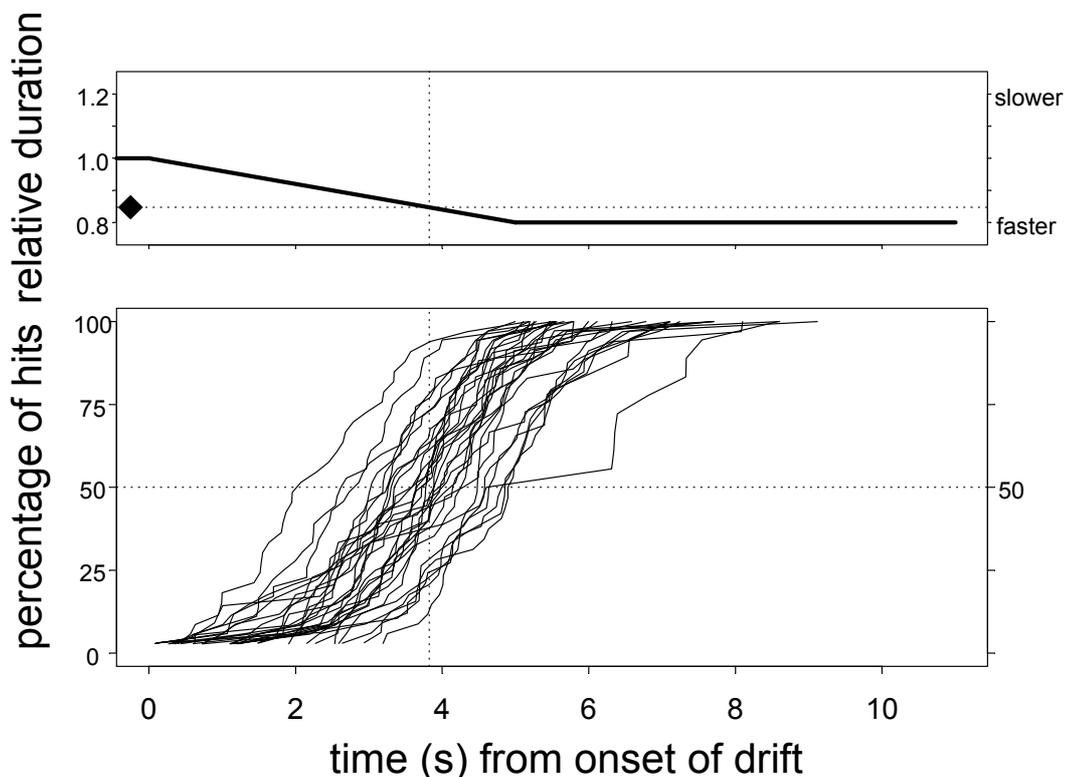
---

<sup>3</sup> Hence the tempo drifts up or down by 4% per second, during 5 seconds.

Each list was presented to 17 listeners, who reported no hearing defects (age mean 26, median 23 years). Listeners were instructed to press a button (with the index finger of their preferred hand) as soon as they detected a change in tempo in the speech stimulus. Response latencies were measured from the onset of tempo drift. Data from 3 participants were lost for various reasons.

## 4.2 Results and discussion

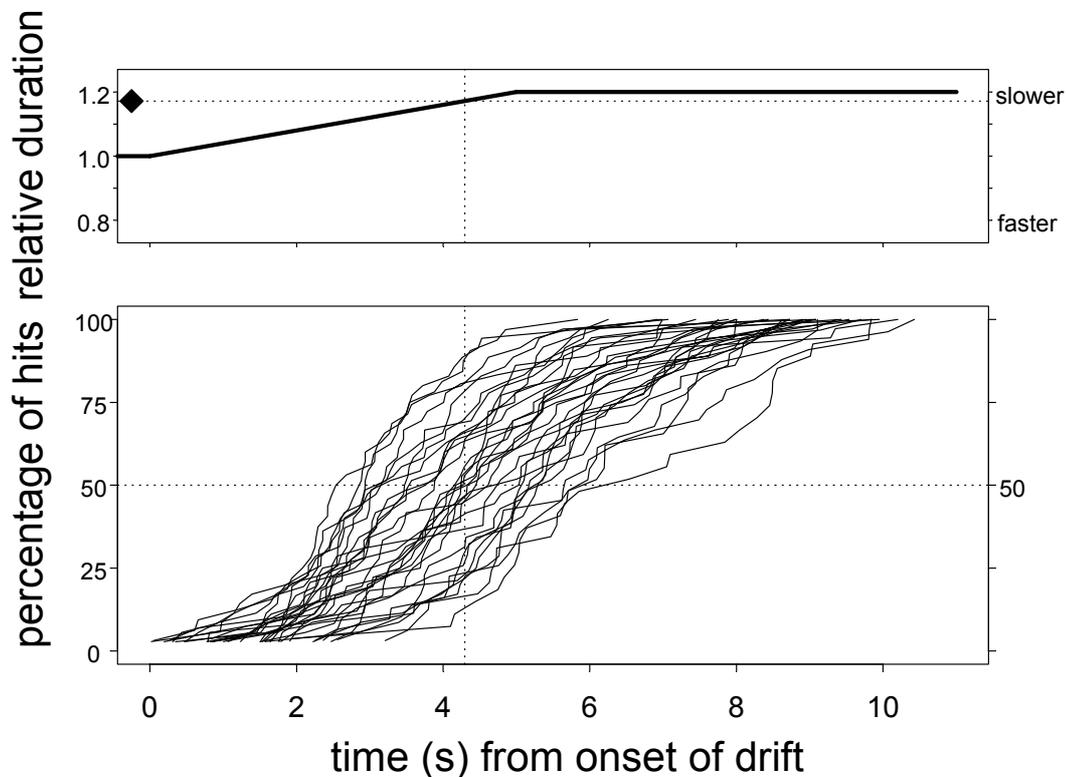
The JND for speech tempo can be obtained from the response latencies, as follows. If a listener responds at 2 s after onset of drift, then he has detected an 8% change in duration or tempo. A listener's JND is defined here as the amount of tempo drift at which he has responded in 50% of the test items. Figures 1 and 2 below show the listeners' proportions of hits, as a function of time after onset of drift. This procedure yields average JNDs of -15.4% for accelerating tempo (relative duration 0.846; Figure 1) and +17.2% for decelerating tempo (relative duration 1.172; Figure 2).



*Figure 1.* Individual listeners' proportions of hits (tempo drift detected) as a function of time (in seconds) after onset of tempo acceleration. The top panel shows the relative duration of the speech stimulus on the same time axis, with the average JND marked.

These results suggest that the average JND for speech tempo across listeners is about 15%, which is considerably larger than the values reported for musical tempo mentioned above. This larger JND value could mean two things. Perhaps the JND for speech tempo is indeed larger than that for music tempo; this could be due to the relatively large perturbatory effects

of articulatory gestures on ‘underlying’ tempo. As an alternative explanation, the method of drift detection used here could be less reliable for speech tempo than for music tempo.



*Figure 2.* Individual listeners’ proportions of hits (tempo drift detected) as a function of time (in seconds) after onset of tempo deceleration. The top panel shows the relative duration of the speech stimulus on the same time axis, with the average JND marked.

Several arguments suggest that this latter explanation may be the more plausible one. First, there is considerable variation among texts or items in their JNDs (for accelerations, range is 8% to 21%, for decelerations 8% to 23%). The tempo manipulation consisted in increasing compression or expansion, over a 5 s time window, irrespective of the contents of that window. This might have increased variability in the obtained JNDs, because speech pauses have not been taken into account. If the manipulation window contained a speech pause of 1 second, from  $t=1$  (where acceleration is 4%) to  $t=2$  (acceleration 8%), then the listener cannot detect intermediate values, because acceleration is not audible within a pause. The onset of the manipulation window was varied, but this variation plus the random presence of pauses within that window may well have increased the noise in listeners’ response times, which increased the resulting JND values<sup>4</sup>.

<sup>4</sup> If listeners did not respond before such pauses, then they could only respond *after* such a pause, and not at an intermediate point in time (during the pause). Hence, the effects described in the text could only have *increased* response times and their JNDs.

Consequently, it was attempted to establish the JND for speech tempo by a more conventional method, in which listeners have to compare two speech stimuli that vary only in the property under investigation, viz. speech tempo.

## 5 Pairwise comparison

### 5.1 Method

Stimulus materials consisted of 20 text passages, selected from the above experiment (5 passages from each speaker). Each was pruned to a fragment that does not contain major pauses (the resulting fragment usually corresponds to a major phrase). Means and standard deviations over the 20 fragments were as follows: duration 3.035 s (0.553), length 8 words (2), length 13.5 syllables (3.4), tempo 159.9 wpm (37.4), average syllable duration 239 ms (77).

These fragments were then accelerated to 0.80, 0.85, 0.87, 0.89, 0.91, 0.93, and 0.95 of the original duration, and decelerated to 1.05, 1.07, 1.09, 1.11, 1.13, 1.15 and 1.20 of the original duration, yielding 7×2 manipulated plus 1 unmanipulated version for each fragment. Temporal compression or expansion was uniform throughout the fragment.

Listeners were 24 students of a PABO college in Utrecht, who reported no hearing defects. They heard two versions of the same passage, with a 600 ms interval between the versions. One version was always the original or reference version, the other was one of the 14 manipulated versions or the original version. Each pair was presented in two orders, with the reference version as either the first or last member of the presentation pair. Listeners' task was to indicate whether the two versions were the same or different (Nootboom & Eefting, 1994). This task was chosen because of its similarity with the detection of tempo drift. The order of the 640 pairs (20 passages × 8 versions × 2 orders × 2 directions) was randomized anew for each listener. Listeners were tested individually in a quiet room. They indicated their same-or-different response by pressing one of two keys, of which the "different" key was always under their dominant hand. Total time of each session was approx. 1.5 hours, including 3 short pauses at regular intervals.

Data from two listeners were discarded: one because of her high miss rate, and the other because of a mild 'cluttering' disorder in his speech. Responses on two manipulated versions that turned out to be defective, were also discarded.

### 5.2 Results and discussion

In a 2IAX design, the JND is sometimes defined as the difference in tempo or duration at which half of the responses are "different". The upper panel of Figure 3 below shows the listeners' percentages of "different" responses, as well as their average. The present experiment also contains a bias, however, which renders the above procedure inappropriate. If both members of a pair are the same, then no "different" responses are expected. Because all versions were compared with an unmanipulated reference version, this situation only occurred with both versions unmanipulated. Interestingly, listeners incorrectly judged these pairs as "different" in 15.1% of their responses; this false-alarm rate deviates significantly from zero (with standard error of 3.7% between listeners' average hit rates,  $p < .001$ ). This possible bias may stem from the 2IAX (same~different) design, which refers to a criterion internal to the

listener (a criterion of equality or difference), which makes this type of experiment susceptible to bias<sup>5</sup>.

Hence, the JND was defined here by means of  $d'$  values, because these are based on the percentages of hits as well as false alarms. The  $d'$  values are plotted in the lower panel of Figure 3. The JND is defined as the difference in tempo or duration at which  $d'=1$ . By this procedure, average JNDs are  $-9.0\%$  for accelerating tempo, and  $+11.5\%$  for decelerating tempo.

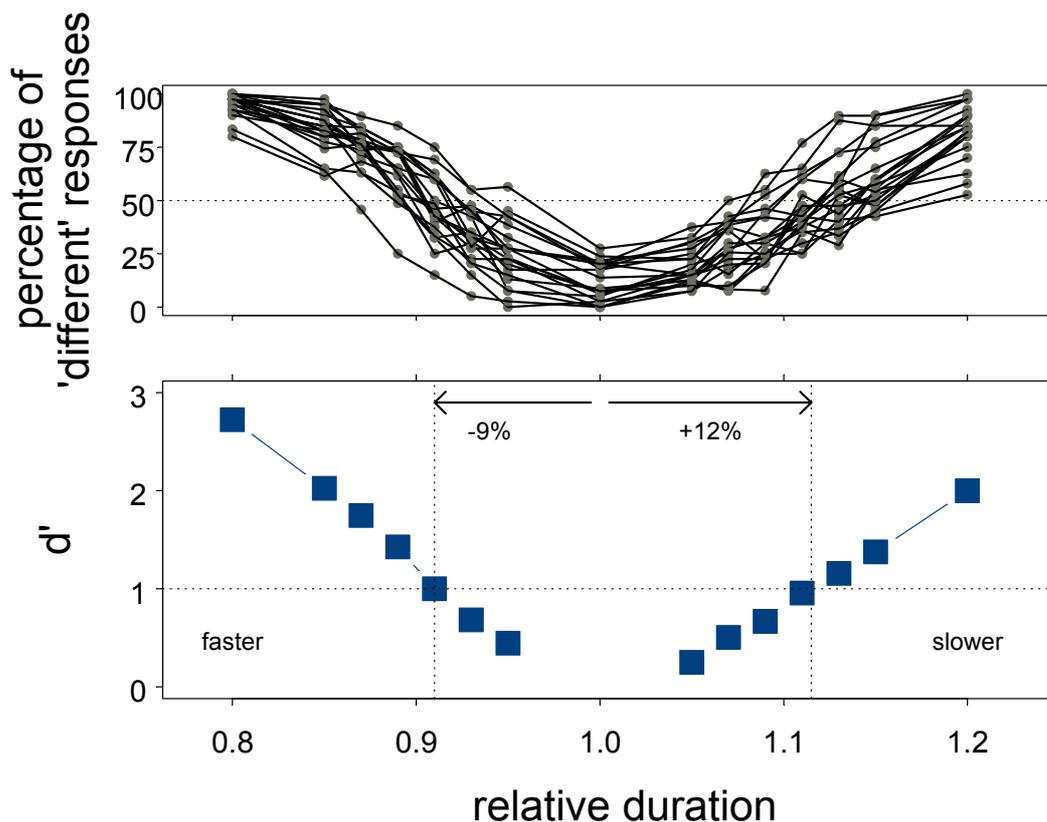


Figure 3. Individual listeners' percentages of "different" responses (*top*), and average values of  $d'$  (*bottom*), as a function of relative duration or tempo.

The lower JNDs obtained in this 2IAX experiment suggest that the method in the previous experiment, viz. detection of tempo drift, is indeed not valid, in that it over-estimates the JND for speech tempo. The present JNDs are more in line with those reported for musical tempo, especially if we allow for the larger variability in speech timing (as compared to musical timing) due to articulatory constraints. A provisional conclusion of this present experiment is that the JND for speech tempo is about 10%.

<sup>5</sup> The 2IFC task does not suffer from this disadvantage: the two versions of a pair are only compared with each other (first-second is faster), without reference to a subjective criterion. This latter method is therefore to be preferred. We hope to re-run the present 2IAX experiment as a 2IFC experiment.

## 6 Discussion

The observed JND of 10% change in speech tempo appears to match the between-speaker distribution of speech tempo in the Spoken Dutch Corpus. One third of the speakers differs by more than 10% from the overall average speaking rate (14 of them are slower than -10%, and another 14 are faster than +10%). Hence, between-speaker differences should be easily noticeable, as indeed they are.

How large do within-speaker variations in speaking rate have to be, in order to be perceptually relevant? The present study suggests that changes in speech tempo smaller than 10% are not noticeable. Hence, the 4% change in tempo reported by Eefting and Nootboom (1993) was indeed not perceptually relevant — in fact, it would be hardly noticeable even in tone sequences or drum machines.

Because some within-speaker changes in tempo *are* noticed, and relevant (see below), these changes must be larger than 10%. A detailed analysis of within-speaker tempo variation in the CGN sample is still pending, but other studies suggest that speakers do indeed accelerate and decelerate by a considerable amount. For example, Nootboom and Eefting (1994) report on a professional speaker, whose average syllable duration per phrase ranged between 118 ms and 289 ms (with mean 176, s.d. 35); this amounts to a change by -33% and +64%. Chafe (2002) reports on a spontaneous conversation in which one speaker accelerates by 33% to convey her high emotional involvement (her average syllable duration decreases from 150 to 100 ms).

Such tempo variation is indeed communicatively relevant, as exemplified by a recent study on the effects of speaking rate on responses to a spoken advertisement, with young adult listeners (Megehee, Dobie, & Grant, 2003). Relative to the original version at normal rate (100%), the time-compressed version (by -15%, to 85% of the original duration) yielded a more positive attitude to the speaker (“trustworthy, secure, favorable”, etc.), and a higher number of affective responses to an open-ended question about the advertised product. The time-expanded version (by +15%, to 115%) yielded a more positive attitude to the message, and a higher number of “cognitive” responses to the open-ended question.

The considerable tempo changes discussed above do not pose great problems for the listener. Speech compressed to 65% of its original duration (by -35%) is still reported to be “perfectly intelligible” (Janse, 2004:160). And if speech is time-compressed even further, to 35% of its original duration (by -65%), which amounts to a highly unnatural speaking rate, even then intelligibility does not fall below 53% correct identifications for real words (Janse, Nootboom, & Quené, 2003).

In conclusion, this study suggests that the just noticeable difference for tempo in speech is about 10%. Further research is necessary, however, to eliminate a possible downward bias in this reported JND. Tempo variations between speakers and within speakers typically exceed this provisional JND, which adds to the importance of these tempo variations in speech communication.

## Acknowledgements

My sincere thanks are due to Grace Postma and Petra van de Ree, for conducting both experiments reported here, to Vincent van Heuven and Sieb Nootboom for helpful comments, to Hans Van de Velde for assistance in the corpus analysis, and to Theo Veenker for technical assistance.

## References

- Benguerel, A.-P., & D'Arcy, J. (1986). Time-warping and the perception of rhythm in speech. *Journal of Phonetics*, 14(2), 231-246.
- Chafe, W. (2002). Prosody and emotion in a sample of real speech. In P. H. Fries, M. Cummings, D. Lockwood & W. Spruiell (Eds.), *Relations and Functions Within and Around Language* (pp. 277-315). London: Continuum.
- Den Os, E. A. (1985). Perception of speech rate of Dutch and Italian utterances. *Phonetica*, 42, 124-134.
- Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54(3), 277-286.
- Eefting, W., & Nootboom, S. G. (1993). Accentuation, information value and word duration: effects on speech production, naturalness and sentence processing. In V. J. Van Heuven & L. C. W. Pols (Eds.), *Analysis and synthesis of speech: Strategic research towards high-quality text-to-speech generation* (pp. 225-240). Berlin: Mouton de Gruyter. Speech Research; 11.
- Eefting, W., & Rietveld, A. C. M. (1989). Just Noticeable Differences of articulation rate at sentence level. *Speech Communication*, 8, 355-361.
- Ellis, M. C. (1991). Thresholds for detecting tempo change. *Psychology of Music*, 19(1), 164-169.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.
- Grondin, S. (2003). *Detection of tempo accelerations and decelerations with abrupt and gradual variations*. Paper presented at the Rhythm Perception and Production Workshop, Île de Tatihou, France.
- Humes, J. C. (2002). *Speak Like Churchill, Stand Like Lincoln: 21 Powerful secrets of history's greatest speakers*. New York: Three Rivers Press.
- Janse, E. (2004). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, 42, 155-173.
- Janse, E., Nootboom, S. G., & Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication*, 41(2-3), 287-301.
- Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58(6), 927-935.
- Lindblom, B. E. F. (1989). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403-439). Dordrecht: Kluwer.
- Megehee, C. M., Dobie, K., & Grant, J. (2003). Time versus pause manipulation in communications directed to the young adult population: Does it matter? *Journal of Advertising Research*, 43(3), 281-292.
- Nootboom, S. G., & Cohen, A. (1984). *Spreken en Verstaan: Een nieuwe inleiding tot de experimentele fonetiek* (2nd ed.). Assen: Van Gorcum.
- Nootboom, S. G., & Eefting, W. (1994). Evidence for the adaptive nature of speech on the phrase level and below. *Phonetica*, 51(1-3), 92-98.
- Oostdijk, N. (2000). Het Corpus Gesproken Nederlands. *Nederlandse Taalkunde* 5 (3), 280-284.
- Perron, M. (1994). *Checking tempo stability of MIDI sequencers*. Paper presented at the 97th Convention of the Audio Engineering Society, San Francisco.
- Pfeuty, M., Ragot, R., & Pouthas, V. (2003). Processes involved in tempo perception: A CNV analysis. *Psychophysiology*, 40(1), 69-76.
- Stetson, R. H. (1988). *Motor Phonetics* (retrospective edition, edited by J.A.S. Kelso & K.G. Munhall). Boston: Little, Brown and Company.
- Van Heuven, V. J. (2003). Vervlakt het Nederlands? Over intonatie. In J. Stroop (Ed.), *Waar gaat het Nederlands naartoe? Panorama van een taal* (pp. 215-223). Amsterdam: Bert Bakker.
- Van Hout, R., De Schutter, G., De Crom, E., Huinck, W., Kloots, H., & Van de Velde, H. (1999). De uitspraak van het Standaard-Nederlands: variatie en varianten in Vlaanderen en Nederland. In E. Huls & B. Weltens (Eds.), *Artikelen van de Derde Sociolinguïstische Conferentie* (pp. 183-196). Delft: Eburon.
- Zwaardemaker, H., & Eijkman, L. P. H. (1928). *Leerboek der Phonetiek: inzonderheid met betrekking tot het Standaard-Nederlandsch*. Haarlem: Erven F. Bohn.