The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners

Versfeld, N.J.; Dreschler, W.A.

DOI
10.1121/1.1426376

Publication date
2002

Published in
The Journal of the Acoustical Society of America

Citation for published version (APA):
Versfeld, N. J., & Dreschler, W. A. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *The Journal of the Acoustical Society of America*, *111*(1), 401-408. https://doi.org/10.1121/1.1426376

The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners

Niek J. Versfeld[a] and Wouter A. Dreschler
Department of Clinical and Experimental Audiology, Academic Medical Center Room D2-330, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

(Received 28 March 2001; revised 10 October 2001; accepted 10 October 2001)

A conventional measure to determine the ability to understand speech in noisy backgrounds is the so-called speech reception threshold (SRT) for sentences. It yields the signal-to-noise ratio (in dB) for which half of the sentences are correctly perceived. The SRT defines to what degree speech must be audible to a listener in order to become just intelligible. There are indications that elderly listeners have greater difficulty in understanding speech in adverse listening conditions than young listeners. This may be partly due to the differences in hearing sensitivity (presbycusis), hence audibility, but other factors, such as temporal acuity, may also play a significant role. A potential measure for the temporal acuity may be the threshold to which speech can be accelerated, or compressed in time. A new test is introduced where the speech rate is varied adaptively. In analogy to the SRT, the time-compression threshold (or TCT) then is defined as the speech rate (expressed in syllables per second) for which half of the sentences are correctly perceived. In experiment I, the TCT test is introduced and normative data are provided. In experiment II, four groups of subjects (young and elderly normal-hearing and hearing-impaired subjects) participated, and the SRT's in stationary and fluctuating speech-shaped noise were determined, as well as the TCT. The results show that the SRT in fluctuating noise and the TCT are highly correlated. All tests indicate that, even after correction for the hearing loss, elderly normal-hearing subjects perform worse than young normal-hearing subjects. The results indicate that the use of the TCT test or the SRT test in fluctuating noise is preferred over the SRT test in stationary noise. [DOI: 10.1121/1.1426376]

PACS numbers: 43.71.Gv, 43.71.Ky, 43.71.Lz [DOS]

I. INTRODUCTION

A conventional measure to determine the ability to understand speech in noisy backgrounds is the so-called speech reception threshold (SRT) for sentences (Plomp and Mimpen, 1979; Nilsson *et al.*, 1994; Versfeld *et al.*, 2000). Typically, simple, meaningful sentences are partially masked by noise, and the signal-to-noise ratio is varied in an adaptive manner such that the critical signal-to-noise ratio is obtained for which 50% of the sentences is completely intelligible. The critical signal-to-noise ratio, or SRT, defines to what degree speech information must be available to the listener in order to make it just intelligible. The SRT test usually utilizes stationary noise to mask the speech signal. It is known that models that operate in the spectral domain, such as the Articulation Index (AI, Fletcher, 1953), and its successor, the Speech Intelligibility Index (SII, ANSI, 1997) are able to make quite accurate predictions for the SRT with various types of noise masker. However, there are indications that especially elderly listeners have more difficulties with understanding speech in adverse listening conditions than young listeners (Konkle *et al.*, 1977; Gordon-Salant and Fitzgibbons, 1993, 1997), even after correction for the hearing loss. Presumably, other factors such as temporal resolution or cognitive demands may also play a significant role (Gordon-Salant and Fitzgibbons, 1993, 1997). Temporal acuity (such as backward masking, cf. Gehr and Sommers, 1999) appears to be most affected by age, in contrast to spectral masking (governed by the auditory filter bandwidth), which seems to remain unaffected as a function of age (Peters and Moore, 1992; Sommers and Gehr, 1998). It is conceivable that the SRT in stationary noise may not be the most appropriate measure to assess the effect of reduced temporal resolution.

An alternative manner to assess temporal acuity in relation to intelligibility is to measure the amount to which speech can be accelerated, or compressed in time. Experiments dealing with time-compressed speech go back to the fifties, where Fairbanks and Kodman (1957) measured word intelligibility as a function of time compression. From there on, time-compressed speech has been used for a variety of topics, including the detection of lesions of the brain stem and auditory cortex (e.g., Beasley *et al.*, 1972a, b; Kurdziel *et al.*, 1976; Beattie, 1986), central auditory processing, both in children (Riensche *et al.*, 1986; Bornstein, 1994; Stollman *et al.*, 1994; Stark *et al.*, 1995) and elderly listeners (Stollman and Kapteyn, 1994; Gordon-Salant and Fitzgibbons, 1997; Vaughan and Letowski, 1997), temporal processing and age effects (Konkle *et al.*, 1997; Gordon-Salant and Fitzgibbons, 1993, 1999), and hearing loss (Kurdziel *et al.*, 1975; Grimes *et al.*, 1984; Stuart and Phillips, 1998). Most recently, time-compressed speech has been used for the assessment of temporal processing in cochlear-implant listeners (Fu *et al.*, 2001). The technique itself has been used to reduce the time needed to listen to a message (Arons, 1992).

---

[a] Electronic mail: N.J.Versfeld@AMC.UvA.nl

With respect to time-compressed speech, a number of topics have not yet been addressed.

First, since both SRT and TCT measure aspects of speech intelligibility, it seems straightforward to study the relationship between these tests. From a theoretical point of view, there is no direct reason to expect SRT and TCT to be highly correlated. Thus, TCT might provide additional information to assess the speech perception capabilities of a given individual if it is not highly correlated to the SRT. It may give more insight into the differences in auditory and cognitive mechanisms involved with these different tests. Alternatively, if TCT and SRT indeed are highly correlated, it seems likely that the same auditory or cognitive mechanisms are underlying. Thus, the purpose of this paper is to assess the relationship between the SRT and TCT. To our knowledge, no paper has yet reported on this relationship (although several papers report on the combined effect of time compression and masking on intelligibility, e.g., Stollman and Kapteyn, 1994; Lacroix and Harris, 1979).

Second, the amount of time compression usually is expressed in a percentage. This does not cause any problems as long as single words are used, uttered by the same speaker. But, it is uncertain whether percentage compression is still a valid measure in the case of different speakers, using different speaking rates. To gain some insight into which measure is perceptually relevant, sentence materials are more appropriate than single-word materials. Unfortunately, only a few papers report on the perception of time-compressed sentences (Vaughan and Letowski, 1997; Gordon-Salant and Fitzgibbons, 1999; Fu *et al.*, 2001).

Third, most certainly due to the availability of the stimulus materials and the computational complexity, intelligibility of time-compressed speech is always measured at fixed time-compression rates (e.g., Beattie, 1986). It is known that fixed-stimuli methods may seriously suffer from ceiling and floor effects, unless the entire psychometric function is measured. Adaptive procedures do not have these problems, so measurements are much more efficient. In the literature, adaptive procedures in combination with time compression have never been reported. Only de Haan and co-workers (de Haan, 1977, 1982; de Haan and Schjelderup, 1978) devised a system with which the speech rate could be varied adaptively. They used this device to assess the relationship between intelligibility and comprehension of speech.

The present paper introduces a test where sentences are time compressed, and an adaptive method is used to define the threshold of intelligibility for time-compressed speech. In analogy to the SRT test, the time-compression threshold (or TCT) is defined as the speech rate (in syllables per second) for which half of the sentences are correctly perceived. In experiment I, normative data for the TCT test are presented for a group of young, normal-hearing subjects. In experiment II, both SRT in either stationary or fluctuating noise and TCT are measured for young and elderly normal-hearing or hearing-impaired listeners. The results of experiment II then are used to assess the relationship between SRT and TCT, and to assess the effect of age and hearing loss upon SRT and TCT.

II. EXPERIMENT I. DEVELOPMENT OF THE TCT TEST AND NORMATIVE DATA

A. Stimuli

Stimuli were 260 meaningful sentences (e.g., ‘‘de bal vloog over de schutting’’ [the ball flew over the fence], ‘‘de voordeur bij de buren klemt’’ [the neighbors’ front door jams]), originally developed for a reliable measurement of the SRT in noise (Plomp and Mimpen, 1979). Each sentence consisted of 4 to 8 words, but always comprised 8 or 9 syllables. Half of this set was uttered by a female speaker (from Plomp and Mimpen, 1979), the other half by a male speaker (from Smoorenburg, 1986). This speech material (without any background noise) was time compressed by means of a modified pitch synchronous overlap add (PSOLA) technique (Moulines and Laroche, 1995). This technique performs duration reduction in the time domain, where it operates in a uniform manner over the entire sentence waveform. PSOLA preserves most physical characteristics of the speech signal, such as, for example, the spectral shape, the periodicity (pitch height), and the amplitude distribution. The PSOLA method has the property that it preserves the naturalness of speech, even at high time-compression rates. Each sentence was time compressed to 11 different degrees. For the male speaker, the speaking rates were $4.576/(0.85)^N$ syllables per second (syll/s), where N ranged between 0 (original speaking rate) and 10 (highest speaking rate). The original speaking rate was determined by manually counting the number of syllables of the entire set, and dividing this number by the total duration in seconds. For the female speaker, the speaking rates were $3.612/(0.85)^N$ syll/s, where again N ranged from 0 to 10. Each sentence was stored into a separate file, resulting in a total of 260 (sentences)$*$11 (speaking rates) $=2860$ files.

B. Subjects

Fourteen young, normal-hearing subjects (4 male, 10 female) participated. Their median age was 22 years and ranged from 20 to 29 years. Their pure-tone thresholds were 15 dB HL or better in their test ear at octave frequencies between 125 and 8000 Hz (inclusive). Subjects were mostly voluntary university students.

C. Procedure

Subjects were tested individually in a sound-insulated booth. Signals were played out via a SoundBlaster soundcard on a PC at a sample frequency of 15 620 Hz, low-pass filtered at 6.5 kHz, and subsequently fed to an InterAcoustics AC40 audiometer. Subjects received the signals via the audiometer’s TDH 39P headphones over their best ear at a fixed level 75 dBA. Three subjects received the stimuli at a level of 60 dBA. (These subjects also participated in another listening experiment, not reported in this paper, that required this stimulus level.) The subject’s task was to repeat the sentence he or she had just been presented. A sentence was scored if the listener repeated every word exactly. The speaking rate was varied adaptively via a one-up, one-down procedure. That is, if a sentence was repeated correctly, the speaking rate of the next sentence was increased. If the sentence was not repeated correctly, the speaking rate of the next sentence was decreased. This procedure makes the speaking rate converge to the point for which half of the (time-compressed) sentences is scored. In analogy to the procedure described by Plomp and Mimpen (1979), 13 sentences formed a list, and one list was required to estimate the threshold for the intelligibility of time-compressed speech, or the time-compression threshold (TCT). The TCT was estimated by taking the geometric mean of the speaking rate of the last ten sentences.

With the existing speech materials, 20 independent lists of 13 sentences each could be formed. Two of these were used as practice lists. The TCT test was interleaved with two SRT tests (reported on in experiment II). Every list occurred at the TCT test, but every subject received only 6 (out of 18) lists for the part dealing with the TCT; the remaining 12 lists were reserved for the SRT part. Lists and sentences within each list were always presented in a fixed order. Conditions (i.e., TCT and SRT), however, were arranged according to balanced design, wherein male and female speaker alternated between lists as well. Within a subject, each sentence was presented only once. The experiment (including the SRT test) lasted about 1 h.

D. Results and discussion

Per subject, 6 TCTs are available, 3 for a male speaker and 3 for a female speaker. In order to provide normative thresholds that are unbiased by learning effects, an analysis of variance (ANOVA) was performed. A three-way (“subject” by “speaker” by “repetition”) ANOVA showed no significant differences between the average threshold obtained in the first list presentation and the second or third list presentation ($F[2,26]=1.58, p>0.1$). Thus, no significant learning effect was observed. Also, a three-way (“repetition” by “speaker” by “level”) ANOVA was performed to assess the effect of level. No significant differences were found between the group of subjects that received the stimuli at presentation levels of 75 and 60 dBA ($F[1,72]=3.26, p>0.05$).

For the present group of 14 normal-hearing subjects, the raw data were pooled. Figure 1 displays the proportion of correct responses as a function of the speaking rate. Open and filled symbols represent results obtained with the male and female speaker, respectively. The solid line is a best fit (in a maximum-likelihood sense) of a logistic function to the data obtained with female speech. Similarly, the dashed line is a best fit for male speech. The speaking rate for which the proportion of correct responses is 0.5 is 12.5 syll/s for the female speaker and 12.8 syll/s for the male speaker. The logistic function is somewhat steeper for the male speaker than for the female speaker, but this difference is not significant ($z=1.1, p>0.1$).

Next, the data per list (i.e., the individual TCTs) were considered. The averages and 95%-confidence intervals of the 14 [subjects]$*$3 [repetitions]$=42$ TCTs are 12.3 syll/s and 1.86 syll/s for the female voice and 12.8 syll/s and 2.05 syll/s for the male voice, respectively. This means that if a TCT test is performed (e.g., in a clinical setting), a TCT lower than 10.5 syll/s is significantly worse than average. The former two ANOVAs show that the small difference between the average TCT of the male and female speaker (12.3 and 12.8 syll/s) is significant ($F[1,26]=5.49$ and $F[1,72]=4.69$, both $p<0.05$). Furthermore, the main effect of subject was significant ($F[13,26]=2.38, p<0.05$), but none of the interactions did reach the 5%-level of significance.

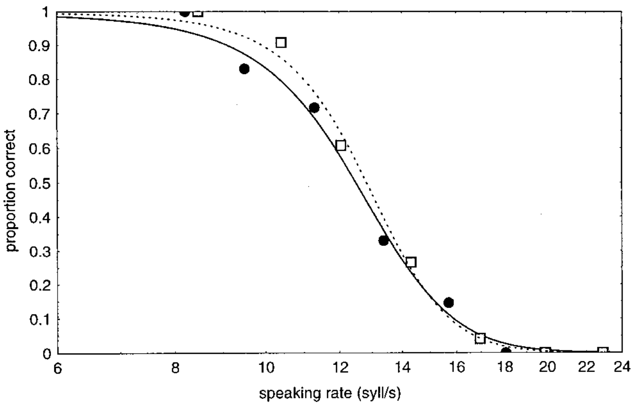

FIG. 1. Proportion of correct responses as a function of speaking rate (syll/s). Open and filled symbols indicate results obtained with the male and female speaker, respectively. The curves represent a best fit of a logistic function to the data.

The slight discrepancy between the two methods (maximum-likelihood method versus TCT averaging) is not only caused by the difference in the method of calculating the threshold, but mostly by the fact that with the TCT averaging only the last ten responses are taken into account (cf. Plomp and Mimpen, 1979), whereas with the maximum-likelihood method all except for the first responses are taken into account.

III. EXPERIMENT II. RELATIONSHIP BETWEEN TCT AND SRT

A. Stimuli

Stimuli were the same 260 sentences (20 lists) as described in experiment I. They were used in both the TCT and SRT test. With the SRT test, the speech was masked by running noise, and the spectrum of this noise was shaped according to the long-term average spectrum of the respective speaker. The noise was either stationary (Plomp and Mimpen, 1979) or fluctuating, resembling the amplitude modulations in a speech signal of a single speaker (cf. Festen and Plomp, 1990).

B. Subjects

In total, 49 subjects participated in this experiment. Fourteen of them also participated in experiment I. All subjects were fluent speakers of the Dutch language. The subjects’ test ear was always their best ear. With this ear they were able to reach at least 80% speech discriminability for monosyllabic words in quiet. All subjects participated on a voluntary basis. According to age and hearing loss, they could be classified into four groups.

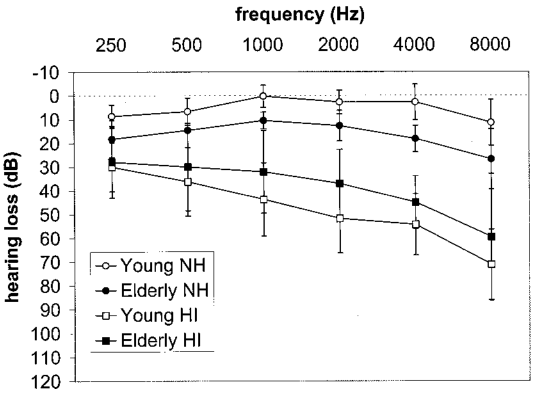

FIG. 2. Audiograms averaged across subjects for the four groups of subjects (see the legend). Error bars denote the standard deviations between subjects.

1. Young, normal-hearing subjects

Eighteen normal-hearing subjects (7 male, 11 female) participated. The median age was 23 years, ranging between 20 and 29 years. Their hearing was on average equal to or better than 10 dB HL at octave frequencies between 250 and 8000 Hz, with extremes up to 35 dB HL at 6000 or 8000 Hz. The average hearing loss is plotted in Fig. 2 as open circles. Error bars indicate the standard deviation between subjects. The TCT data of 14 subjects in this group have been used in experiment I.

2. Elderly, normal-hearing subjects

Eleven elderly, normal-hearing subjects (2 male, 9 female) participated. Their median age was 64 years and ranged from 58 to 70 years. Their pure-tone thresholds were 30 dB HL or better in their test ear at octave frequencies between 250 and 6000 Hz, with extreme values up to 60 dB HL at 8000 Hz. The average pure-tone threshold (in dB HL) is given in Fig. 2 as filled circles. Error bars indicate the standard deviation between subjects.

3. Young, hearing-impaired subjects

Eight young, hearing-impaired subjects (5 male and 3 female ranging in age from 15 to 35, median age 18 years) participated. Most of them were recruited from a local school for the hearing impaired. All suffered from sensorineural hearing loss (of which five were of hereditary origin, and three were of unknown origin), and their average pure-tone thresholds (in dB HL) are plotted in Fig. 2 as open squares. Error bars indicate the standard deviation between subjects. Apart from the hearing impairment, one subject suffered from additional speech and language problems, and for five subjects, parents were non-native speakers.

4. Elderly, hearing-impaired subjects

Twelve elderly, hearing-impaired subjects (6 male, 6 female) participated. Their median age was 63 years, ranging from 55 to 73 years. They suffered from mild-to-moderate (sensorineural) hearing loss, which all of them had acquired at later age. Their average pure-tone hearing loss is plotted in Fig. 2 as filled squares. Error bars indicate the standard deviation between subjects.

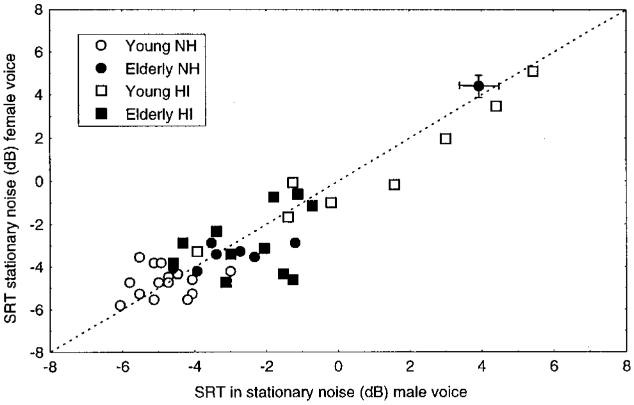

FIG. 3. SRT (dB) in stationary noise. Results obtained with the female speaker plotted as a function of results for the male speaker. The dashed line represents the points of equal SRT.

C. Procedure

With respect to the part dealing with the TCT test, the experimental procedure was that of experiment I. However, speech was presented at a minimum level of 60 dBA, and was at least 20 dB above threshold in quiet. For the individual subject, the masking noise of the SRT test was kept fixed, and was presented at a level equal to the speech in the TCT test. Speech-to-noise ratio was varied by variation of the speech level. The procedure with the SRT test was that described by Plomp and Mimpen (1979). Each list comprised 13 sentences. The subject's task was to reproduce the sentence he or she had just been presented. A sentence was scored if the listener repeated every word exactly. The speech-to-noise ratio was varied adaptively via a one-up, one-down procedure, converging to that speech-to-noise ratio (the speech reception threshold or SRT) for which half of the sentences are scored. The SRT was estimated by averaging the speech-to-noise ratio after the last ten sentences.

Lists and sentences within each list were presented in a fixed order. Male and female speaker alternated between lists. Conditions alternated in a balanced order, and were counterbalanced between subjects. Across subjects and within each subject, each condition and speaker occurred equally often. Within a subject, each sentence was presented only once. Therefore, apart from two practice lists, a subject received six lists with time-compressed speech, six lists with speech in stationary noise, and six lists with speech in fluctuating noise. The experiment lasted about 1 h.

D. Results and discussion

First, correlations between the results obtained for the different tests (i.e., SRT, TCT, male or female speaker) will be discussed. Second, results obtained for different subject groups will be compared. As in experiment I, no significant learning effects were observed, not even for the elderly group or hearing-impaired group.

1. Relationship between male and female speech

Figures 3–5 display the relationship between the results obtained with the male and female speaker, in the case of

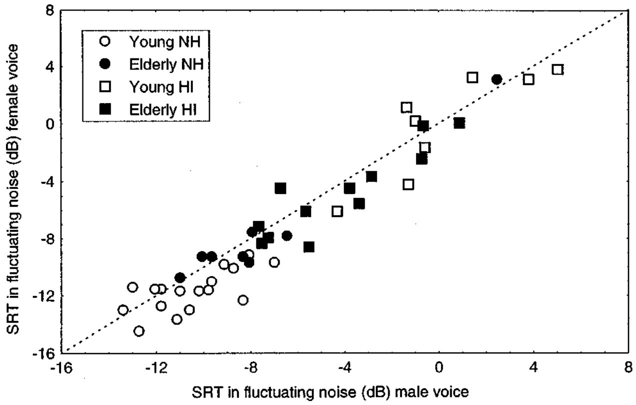

FIG. 4. As in Fig. 3, but for the SRT (dB) in fluctuating noise.

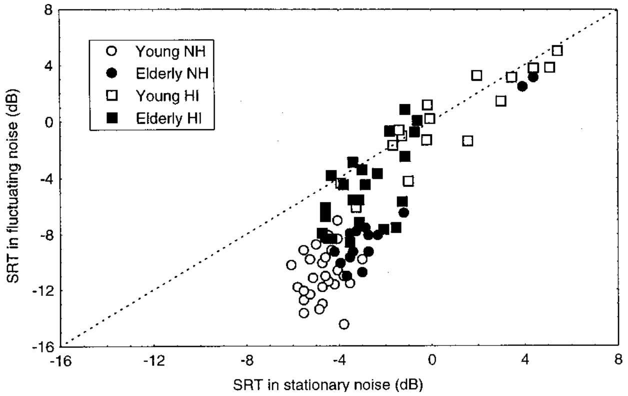

FIG. 6. SRT in fluctuating noise as a function of the SRT in stationary noise. Different symbols indicate the different groups of subjects (see the legend). The dashed line indicates the points of equal SRT.

SRT for stationary noise (hereafter denoted as $\mathrm{SRT_S}$), SRT for fluctuating noise (hereafter denoted as $\mathrm{SRT_F}$), and TCT, respectively. Figure 3 shows the results obtained for speech in stationary noise. The SRT obtained with the female speaker is plotted as a function of the SRT obtained with the male speaker. Different symbols indicate the four different subject groups. Each symbol represents the average of three SRT estimates. The standard error of this estimate is on average 0.5 dB for both speakers, and is indicated with error bars in one of the symbols to the top right of the figure. The dashed curve is a straight line indicating equal SRT. As can be seen from Fig. 3, the deviation from the symbols to the dashed line is small, but there seems to be a slight tendency for the female speaker to yield better results. Indeed, linear regression shows that the best-fitted straight line with unity slope significantly deviates from the dashed line, and the offset is 0.15 dB. Figure 4 shows similar results for fluctuating noise. Here, the offset is 0.65 dB. With time-compressed speech, thresholds obtained with male and female speech are also slightly different, as can be seen in Fig. 5. In contrast, performance for the female speech here is worse, and the difference between male and female speech is 0.55 syll/s. Correlation coefficients between the scores of the male and female speaker were 0.91 for the $\mathrm{SRT_S}$, 0.96 for the $\mathrm{SRT_F}$, and 0.92 for the TCT. In conclusion, differences between male and female speaker were significant, but small. In this paper, all results have been analyzed for the male and female speaker separately, as well as for the pooled data. In none of the cases were results significantly different, so in the remainder of this paper only the pooled data will be presented.

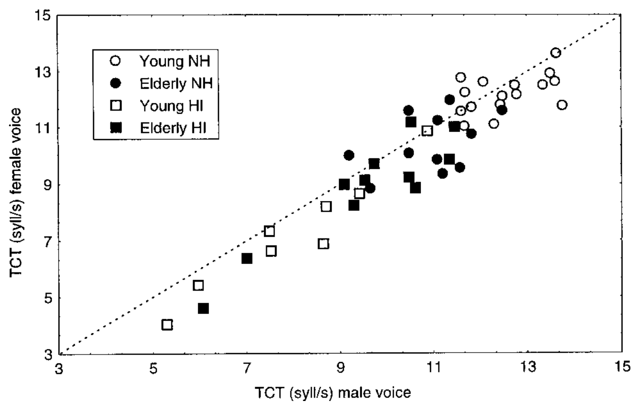

FIG. 5. As in Fig. 3, but for the TCT (syll/s).

2. Relationship between $\mathrm{SRT_S}$, $\mathrm{SRT_F}$, and TCT

Figure 6 displays the relationship between $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$. The dashed line indicates the points where $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$ are equal. For young, normal-hearing listeners, the $\mathrm{SRT_S}$ is on average −4.7 dB, whereas the $\mathrm{SRT_F}$ is much better, namely −11.1 dB. Note that for subjects with a poor $\mathrm{SRT_F}$, $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$ are about equal, whereas a relatively good performance for the $\mathrm{SRT_S}$ does not imply good performance for the $\mathrm{SRT_F}$.

Figure 7 displays the relationship between $\mathrm{SRT_S}$ and TCT. As SRT increases, TCT decreases, as to be expected. However, especially at higher SRTs, there is a considerable amount of scatter in the data. The correlation between the two data sets is 0.73. The correlation between $\mathrm{SRT_F}$ and TCT is considerably higher, namely 0.87. This relationship is depicted in Fig. 8.

3. Effect of age and hearing impairment on SRT and TCT

Table I displays for each subgroup the $\mathrm{SRT_S}$, $\mathrm{SRT_F}$, and TCT. (The values in parentheses are SII values and will be discussed below.) $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$ for the group of young and

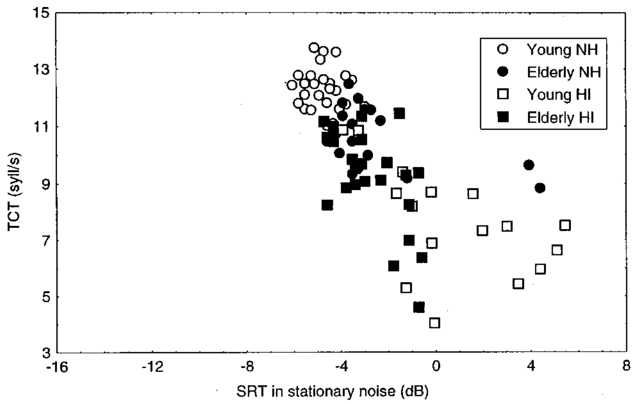

FIG. 7. TCT as a function of the SRT in stationary noise. Different symbols indicate the different groups of subjects (see the legend).

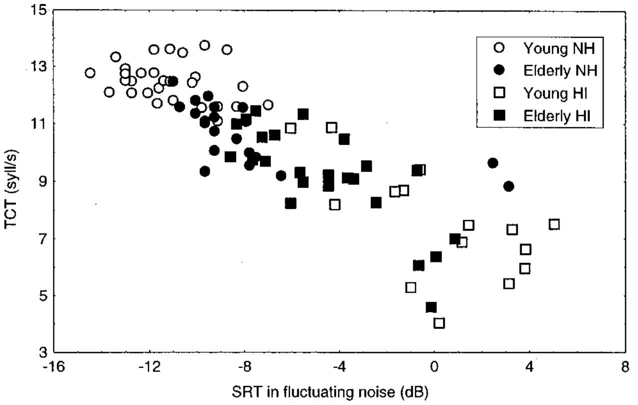

FIG. 8. TCT as a function of the SRT in fluctuating noise. Different symbols indicate the different groups of subjects (see the legend).

elderly normal-hearing subjects are in agreement with those reported in the literature (Festen and Plomp, 1990), and the difference between these two thresholds is about 6.4 dB. With elderly normal-hearing subjects, thresholds are somewhat poorer, but the difference between $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$ still is 5.2 dB. The remarkable difference between the normal-hearing group and the hearing-impaired group is that the difference between $\mathrm{SRT_F}$ and $\mathrm{SRT_S}$ has practically vanished: 0.7 and 1.9 dB for the young and elderly group, respectively. Note that the $\mathrm{SRT_S}$ for both elderly groups is similar.

IV. GENERAL DISCUSSION

The results of experiment I showed that if the speaking rate was expressed in syllables per second, the psychometric functions for the male and female speaker practically overlap, and the difference is 0.3 syll/s. If one keeps in mind that the original speaking rate for the female speaker is on average 3.612 syll/s and for the male speaker 4.576 syll/s (ratio of 1.27), the finding that the two curves overlap to a high degree (ratio equal to 12.8/12.5=1.02) indicates that speaking rate expressed in syllables per second is a more perceptually relevant measure than percentage compression. Unfortunately, no distinction could be made between syllables per second and phonemes per second, since these two measures are highly correlated. However, the source of difficulty with speeded speech seems likely to occur at the phonemic level, with the possibility that consonant compression influences recognition performance more so than vowel compression or overall changes in sentence duration. Other factors, like rate of information processing, may be important as well.

In experiment II, it was observed that the difference between SRT for stationary and fluctuating noise was large for normal-hearing subjects and small for hearing-impaired subjects. This is in accordance with the results of Festen and Plomp (1990), and the difference in SRT is attributed to the notion that normal-hearing listeners are able to use the speech information at the time intervals when the level of the fluctuating masking noise is low. Due to loss of temporal acuity, hearing-impaired subjects often are not able to do so, causing the fluctuating noise to become smeared, behaving like stationary noise, such that the $\mathrm{SRT_S}$ and $\mathrm{SRT_F}$ are about equal. This is exactly what can be seen in Fig. 6. Subjects that perform well in fluctuating noise (low $\mathrm{SRT_F}$) also perform well in stationary noise (low $\mathrm{SRT_S}$), whereas the opposite is not true. This finding leads to the recommendation to conduct the SRT in fluctuating noise only, since it provides more information.

TABLE I. Group-averaged results. For each group the mean SRT (in dB) and TCT (in syll/s) is given. In parentheses the corresponding SII value is given (see the text).

	$\mathrm{SRT_S}$ (dB)	$\mathrm{SRT_F}$ (dB)	TCT (syll/s)
Young, normal hearing	−4.8 (0.300)	−11.1 (0.089)	12.4 (0.321)
Elderly, normal hearing	−2.6 (0.365)	−7.8 (0.188)	10.7 (0.376)
Young, hearing impaired	0.8 (0.374)	0.1 (0.361)	7.6 (0.358)
Elderly, hearing impaired	−2.7 (0.308)	−4.6 (0.239)	9.2 (0.308)

The finding that TCT correlates better with the SRT in fluctuating noise ($r=0.87$) than with the SRT in stationary noise ($r=0.73$) suggests that factors dealing with temporal processing play a dominant role. However, correlation is not as high as the correlation between male and female speech (r between 0.91 and 0.96). So, next to temporal factors, other factors may play a role. One factor may be cognition, for an increase in speech rate implies an increase in information rate which, in turn, requires a larger processing capacity. Thus, the TCT test also may provide information about the cognitive abilities. Especially in elderly subjects, processing capacities are said to decline (e.g., Gordon-Salant and Fitzgibbons, 1997; Salthouse, 1996). Unfortunately, the present data cannot be used to support this assertion, because differences in correlation are only small.

TCT and SRT in fluctuating noise correlate highly. Although time-compressed speech is less natural, it has the advantage that the test is far more time efficient. The duration of one TCT list was on average 83 s, whereas it was 113 and 120 s for the SRT test in stationary and fluctuating noise, respectively. Moreover, TCT may prove to be more successful with children, since the speech sounds funny and makes the task more challenging. However, TCT (and its relationship to SRT) for children has yet not been investigated.

To assess the effect of age on SRT and TCT, a two-way (age [2]×hearing status [2]) ANOVA could be performed. However, the interpretation of the results of this ANOVA are contaminated by the fact that, especially for the young, hearing-impaired group, additional speech and language problems must have played a role, due to their hearing loss on the one hand, and, for five of the subjects, their non-native Dutch-speaking parents on the other hand. Indeed, the data show this poor performance. Therefore, scores between normal-hearing and hearing-impaired subjects cannot be compared. Also, normal-hearing young and elderly subjects cannot be directly compared, since the hearing loss for the elderly group was on average about 10 dB higher than for the young, normal-hearing group. To account for these differences in hearing loss, for every individual subject and condition, scores were converted to Speech Intelligibility Indices (SIIs, ANSI, 1997). For the sake of completeness, the data of all subjects were transformed.

Speech intelligibility index (SII)

The Speech Intelligibility Index (ANSI, 1997) is a calculation scheme that determines the part of the speech spectrum that is audible, i.e., is not below the absolute hearing level and not masked by interfering noise. For each frequency band (usually one-third-octave bands) the proportion of audible speech is calculated and these proportions are weighted summed, since not every band is equally important for speech intelligibility. This results in a number between zero and unity. Completely inaudible speech results in an SII of zero, completely audible speech in an SII of unity.

By definition, at the SRT, 50% of the sentences are perceived correctly. This critical speech-to-noise ratio can be converted to an SII value. Thresholds for subjects that need only little information for speech reception will yield low SII values. Thresholds for subjects that have problems with speech reception, on the other hand, will yield larger SII values. The advantage of using SII instead of SRT is that the SII takes into account the differences in threshold. Thus, it is quite conceivable that a hearing-impaired subject produces high SRTs but low SIIs, whereas a normal-hearing subject with poor speech processing capabilities produces high SIIs.

Table I displays, in parentheses, the SII group-averaged values for the four groups of subjects for the three conditions. SII values for SRT_S are between 0.3 and 0.4, indicating that roughly one-third of all speech information is necessary to reach the threshold of speech intelligibility with short, everyday sentences. Table I shows that, even after correction for the hearing loss, elderly, normal-hearing subjects perform worse than young, normal-hearing subjects ($F[1,27]=104, p<0.005$). Young, hearing-impaired subjects perform even worse, probably due to language problems. On the other hand, in contrast to the SRT_S values, elderly, hearing-impaired subjects perform rather well, reaching SII scores close to the young, normal-hearing group. It seems as if this group is very efficient in the use of speech information.

Unfortunately, SII is a purely spectral measure, i.e., it determines the part of the speech spectrum that exceeds the masking noise. It does not take into account the temporal characteristics of the noise. Therefore, thresholds obtained with time-compressed speech cannot be converted automatically to SII values, since the SII does not depend on the amount of time compression (because the noise spectrum and the speech spectrum do not depend on the amount of time compression). Yet, in order to be able to convert the TCT values to SII values, a simple addition was made to the SII calculation scheme. The SII for the unprocessed condition was calculated and simply multiplied with the amount of time compression, i.e., the sentence duration of the processed speech divided by the original duration. Thus, for example, if the duration of the time-compressed sentence was 0.6 times the original duration, the SII value also was multiplied with a factor 0.6. This calculation scheme can also be interpreted as removing the part of the information that has been removed by cutting out portions of the waveform. The group-averaged results of the SII calculations for the TCTs are given in Table I. The absolute values are very close to the ones corresponding to the SRT_S, and similar trends are even observed.

The SII model was not really developed for speech in fluctuating noise. Nonetheless, SRT_Fs were converted to SIIs, assuming it was stationary noise. The results are given in Table I. The absolute values do not have any meaning, but the relative values show that especially the young, hearing-impaired group has extreme difficulties in taking advantage of the fluctuating nature of the masking noise. Again, both groups of normal-hearing subjects benefit from the fluctuating characteristics, more than the elderly group of hearing-impaired subjects (who did perform well in stationary noise).

V. CONCLUSIONS

Two experiments with time-compressed speech have been reported. The first experiment provided normative data for time-compressed sentences. Speaking rate expressed in syllables per second appeared to be a more perceptually relevant measure than, e.g., compression rate. Young, normal-hearing subjects were able to understand 50% of the sentences at speaking rates of 12.5 syll/s.

The second experiment dealt with the relationship between the threshold for time-compressed speech (TCT) and speech-in-noise (SRT). Both stationary and fluctuating noise were used as a masker. Both young- and elderly normal-hearing and hearing-impaired subjects participated. The results show that (1) TCT correlates better with the SRT in fluctuating noise than with SRT in stationary noise, suggesting that factors dealing with temporal processing play a dominant role in the TCT; (2) SRT in fluctuating noise (and hence TCT) provides more information about the remaining hearing capacity than SRT in stationary noise; (3) even after correction for differences in hearing loss (by means of the SII), elderly subjects perform worse than young subjects in the normal-hearing population; (4) elderly, hearing-impaired subjects perform on average as well as young, normal-hearing subjects in stationary masking noise, but have considerably more difficulties in fluctuating noise.

With respect to audiological testing, the results show that SRT in fluctuating noise is recommended over SRT in stationary noise. One may even consider using TCT due to its time efficiency, despite the fact that time-compressed speech sounds less natural.

ACKNOWLEDGMENTS

The duration manipulation of the sentences was done with the software package GIPOS, developed by Ercan Gigi at the Institute for Perception Research (IPO), Eindhoven, the Netherlands. The first author acknowledges IPO for its hospitality and Ercan for his kind assistance. Marscha Krone, Bas Franck, and especially Maite Tromp are acknowledged for their part in the preparation and conduction of the experiment.

ANSI (**1997**). ANSI S3.5-1997, ''American national standard methods for the calculation of the speech intelligibility index'' (American National Standards Institute, New York).

Arons, B. (**1992**). ''Techniques, perception, and applications of time-compressed speech,'' Proceedings of 1992 Conference, American Voice I/O Society, 169–177.

Beasley, D. S., Forman, B. S., and Rintelmann, W. F. (**1972a**). ''Perception of time-compressed CNC monosyllables by normal listeners,'' J. Aud Res. **12**, 71–75.

Beasley, D. S., Schwimmer, S., and Rintelmann, W. F. (**1972b**). ''Intelligibility of time-compressed CNC monosyllables,'' J. Speech Hear. Res. **15**, 340–350.

Beattie, R. C. (**1986**). ''Normal intelligibility functions for the Auditec CID W-22 test at 30% and 60% time-compression,'' Am. J. Otol. **7**, 60–64.

Bornstein, S. P. (**1994**). ''Time compression and release from masking in adults and children,'' J. Am. Acad. Audiol. **5**, 89–98.

de Haan, H. (**1977**). ''A speech-rate intelligibility threshold for speeded speech and time-compressed connected speech,'' Percept. Psychophys. **22**, 366–372.

de Haan, H. (**1982**). ''The relationship of estimated comprehensibility to the rate of connected speech,'' Percept. Psychophys. **32**, 27–31.

de Haan, H. J., and Schjelderup, J. R. (**1978**). ''Threshold of intelligibility/comprehensibility of rapid connected speech: Method and instrumentation,'' Behav. Res. Methods Instrum. **10**, 841–844.

Fairbanks, G., and Kodman, F. (**1957**). ''Word intelligibility as a function of time compression,'' J. Acoust. Soc. Am. **29**, 636–644.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Fu, Q.-J., Galvin III, J. J., and Wang, X. (**2001**). ''Recognition of time-distorted sentences by normal-hearing and cochlear-implant listeners,'' J. Acoust. Soc. Am. **109**, 379–384.

Gehr, S. E., and Sommers, M. S. (**1999**). ''Age differences in backward masking,'' J. Acoust. Soc. Am. **106**, 2793–2799.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). ''Temporal factors and speech recognition performance in young and elderly listeners,'' J. Speech Hear. Res. **36**, 1276–1285.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). ''Selected cognitive factors and speech recognition performance among young and elderly listeners,'' J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). ''Profile of auditory temporal processing in older listeners,'' J. Speech Lang. Hear. Res. **42**, 300–311.

Grimes, A. M., Mueller, H. G., and Williams, D. L. (**1984**). ''Clinical considerations in the use of time-compressed speech,'' Ear Hear. **5**, 114–117.

Konkle, D. F., Beasley, D. S., and Bess, F. H. (**1977**). ''Intelligibility of time-altered speech in relation to chronological aging,'' J. Speech Hear. Res. **20**, 108–115.

Kurdziel, S. A., Noffsinger, D., and Olsen, W. (**1976**). ''Performance by cortical lesion patients on 40% and 60% time-compressed materials,'' J. Am. Audiol Soc. **2**, 3–7.

Kurdziel, S. A., Rintelmann, W. F., and Beasley, D. S. (**1975**). ''Performance of noise-induced hearing-impaired listeners on time-compressed consonant–nucleus–consonant monosyllables,'' J. Am. Audiol Soc. **1**, 54–60.

Lacroix, P. G., and Harris, J. D. (**1979**). ''Multiplicative effects on sentence comprehension for combined acoustic distortions,'' J. Speech Hear. Res. **22**, 259–269.

Moulines, E., and Laroche, J. (**1995**). ''Non-parametric techniques for pitch-scale and time-scale modification of speech,'' Speech Commun. **16**, 175–205.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.

Peters, R. W., and Moore, B. C. J. (**1992**). ''Auditory filter shapes at low frequencies in young and elderly hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 256–266.

Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Riensche, L. L., Curran, C. E., and Porch, B. E. (**1986**). ''The assessment of reading readiness using multi-dimensionally scored time-compressed speech,'' J. Aud. Res. **26**, 1–4.

Salthouse, T. A. (**1996**). ''The processing-speed theory of adult age differences in cognition,'' Psychol. Rev. **103**, 403–428.

Smoorenburg, G. F. (**1986**). ''Speech reception in individuals with noise-induced hearing loss and its implication for hearing loss criteria,'' in *Basic and Applied Aspects of Noise-induced Hearing Loss*, edited R. J. Salvi, D. Henderson, R. P. Hamernik (Plenum, New York).

Sommers, M. S., and Gehr, S. E. (**1988**). ''Auditory suppression and frequency selectivity in older and younger adults,'' J. Acoust. Soc. Am. **103**, 1067–1074.

Stark, R. E., and Montgomery, J. W. (**1995**). ''Sentence processing in language-impaired children under conditions of filtering and time compression,'' Appl. Psycholing. **16**, 137–154.

Stollman, M. H. P., and Kapteyn, T. S. (**1994**). ''Effect of time-scale modification of speech on the speech recognition threshold in noise for elderly listeners,'' Audiology **33**, 280–290.

Stollman, M. H. P., Kapteyn, T. S., and Wegener Sleeswijk, B. (**1994**). ''Effect of time-scale modification of speech on the speech recognition threshold in noise for hearing-impaired and language-impaired children,'' Scand. Audiol. **23**, 39–46.

Stuart, A., and Phillips, D. P. (**1998**). ''Recognition of temporally distorted words by listeners with and without a simulated hearing loss,'' J. Am. Acad. Audiol **9**, 199–208.

Vaughan, N. E., and Letowski, T. (**1997**). ''Effects of age, speech rate, and type of test on temporal auditory processing,'' J. Speech Lang. Hear. Res. **40**, 1192–1200.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). ''Method for the selection of sentence materials for efficient measurement of the speech reception threshold,'' J. Acoust. Soc. Am. **107**, 1671–1684.