UvA-DARE (Digital Academic Repository)

Release from informational masking by time reversal of native and non-native interfering speech

Rhebergen, K.S.; Versfeld, N.J.; Dreschler, W.A.

DOI

[10.1121/1.2000751](https://doi.org/10.1121/1.2000751)

Publication date

2005

Published in

The Journal of the Acoustical Society of America

[Link to publication](#)

Citation for published version (APA):

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, 118(3 pt 1), 1274-1277. <https://doi.org/10.1121/1.2000751>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Release from informational masking by time reversal of native and non-native interfering speech (L)

Koenraad S. Rhebergen^{a)}

Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-223, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Niek J. Versfeld^{b)}

Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-330, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Wouter. A. Dreschler^{c)}

Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-240, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

(Received 8 December 2004; revised 17 June 2005; accepted 17 June 2005)

In many studies, the influence of intelligibility of the interfering speech is avoided by reversing it in time. Usually, intelligibility with time-reversed interfering speech indeed is higher compared to that with normal interfering speech. However, due to the nature of speech, reversed speech also gives rise to increased forward masking. The latter will result in a decrease in intelligibility. Thus, differences in intelligibility as a consequence of reversing speech in time are due to two opposite effects. This paper describes a speech reception threshold (SRT) test with intelligible and unintelligible interfering speech played normally and time-reversed. With Dutch listeners, Swedish reversed interfering speech gave a rise in SRT of 2.3 dB compared with the Swedish interfering speech (played normally). The difference can be attributed to differences in forward masking. Dutch time-reversed interfering speech gave a decrease in SRT of 4.3 dB compared to (intelligible) Dutch interfering speech. The latter is the result of both a release from informational masking and an increase in forward masking. Therefore, the amount of informational masking is larger than 4.3 dB and, if one assumes similar differences in forward masking for Dutch and Swedish speech, may amount to 6.6 dB. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2000751]

PACS number(s): 43.71.Gv, 43.66.Mk, 43.66.Dc, 43.71.An [KWG]

Pages: 1274–1277

I. INTRODUCTION

There are various papers describing speech intelligibility in the presence of one or more interfering talkers (e.g., Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Drullman and Bronkhorst, 2000; Brungart, 2001; Brungart *et al.*, 2001; Brungart and Simpson, 2002; Summers and Molis, 2004). Remarkably, intelligibility appears to vary greatly between conditions, which in part might be due to the degree of similarity between the target speaker (here: “target” or “signal”) and the interfering speakers (here: “masker” or “noise”). The more similar the target and masker are, the more the listener is confused or distracted, which in turn results in poorer performance. Differences between target and masker with respect to gender (male or female speech) or intelligibility of the interfering speech (native or foreign language) have only a small effect with respect to actual energetic masking of the target, but certainly can have a large effect on intelligibility. The phenomenon of excess masking is often labeled as informational masking (Drullman and Bronkhorst, 2000; Bronkhorst, 2000;

Brungart, 2001; Brungart *et al.*, 2001). This paper concentrates on the part of informational masking due to intelligibility of the interfering speech.

In order to study the speech intelligibility in speech-like maskers, and at the same time avoid the informational masking component due to the intelligibility of the interfering speech, many researchers use fluctuating speech-shaped noise (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Peters *et al.*, 1998; Versfeld and Dreschler, 2002) or play the interfering speech signal backwards (Festen and Plomp, 1990; Summers and Molis, 2004). Fluctuating speech-shaped noise is made by modulating the long term speech spectrum with the speech envelope of the interfering talker. The envelope can be extracted from broad band speech (one band), two bands (Festen and Plomp, 1990), three bands (ICRA noise; Dreschler *et al.*, 2001), or even more. The fluctuating speech-shaped noise has more or less the same intensity fluctuations in time as real speech, has the same long-term spectrum of speech, but is generally unintelligible because the fine structure is lost.

By using time-reversed speech as a masker, the spectral contents of speech are in essence untouched. The reversed speech is not intelligible and therefore this component of the informational masking by definition is removed. However, the envelope of time-reversed speech is reversed as well. The shape of temporal envelope of speech is typically dominated

^{a)}Electronic mail: k.s.rhebergen@amc.uva.nl

^{b)}Electronic mail: n.j.versfeld@amc.uva.nl

^{c)}Electronic mail: w.a.dreschler@amc.uva.nl

by plosive sounds (Rosen, 1992), and the envelope of these sounds is characterized by quick onsets (steep slope) and slow decays (shallow slope). Reversal of the speech signal thus results in temporal envelopes that have abrupt offsets. Since the auditory system does not follow the offset instantaneously, but rather displays a decay of the envelope across time, abrupt changes cannot be followed accurately, hence soft signals can be easily masked by a preceding strong signal (called forward masking).

Irino and Patterson (1996), Carlyon (1996), Stecker and Hafter (2000), and Schlauch *et al.* (2001) examined the perception of stimuli with ramped envelopes (gradual attack and abrupt decay) and damped envelopes (abrupt attack and gradual decay). The sum of these studies reveal that ramped signals are subjectively judged longer in duration and are perceived louder compared to damped signals. This is explained by the abrupt offset at a high level of a ramped sound which results in a persistence of perception, and more forward masking compared to a damped sound. The same effects could be expected by time reversed speech since its envelope is more ramped compared to the damped normally played speech (Rosen, 1992).

Forward masking has a clear effect on speech intelligibility (Festen, 1987). The recovery time from forward masking is in the range of 100–200 ms (Moore, 2003). For completeness, a second type of masking, present in human auditory perception, should be mentioned, viz., backward masking. In this case, a soft signal is masked by a louder signal that follows it. The phenomenon of backward masking is still poorly understood (Moore, 2003). The amount of backward masking obtained with practiced subjects often is little or none (Moore, 2003). Its effect on speech intelligibility is still unclear and probably not very large.

By reversal of the speech masker in time, one expects on the one hand an improvement in intelligibility of the target speech due to the fact that the masking speech becomes unintelligible (partly release from informational masking), and, on the other hand a decrease in performance due to the fact that the temporal envelope of the reversed speech causes an increase of forward masking. Festen and Plomp (1990) and Summers and Molis (2004) found a slight improvement in performance when using reversed speech instead of normal speech as a masker. This improvement is the result of both effects combined. From these experiments it is not possible to assess the individual contribution of both components.

The goal of the current experiment is to separate informational masking due to intelligibility of the interfering speech from forward masking. Additional masking due to time reversal of the speech is assessed by measuring the speech intelligibility in masking speech of a foreign language. Since the foreign speech is not intelligible (neither forward nor backward), differences in intelligibility due to time reversal must be due to differences in forward masking alone. The same experiment is repeated in native interfering speech. Differences in intelligibility due to time reversal then are the combined effect of an increase in forward masking and a decrease in informational masking. Since the speech materials are similar, and since the single effect of additional

masking due to time reversal is known, the effect of informational masking due to intelligibility of the interfering speech can be estimated.

II. METHOD

A. Subjects

Eight normal-hearing subjects (3 male, 5 female) participated. Their mean age was 25 years and ranged from 21 to 39 years. Subjects were native speakers of the Dutch language. Subjects had at least high school education. Each subject had pure-tone thresholds of 15 dB HL or better at octave frequencies from 125 to 8000 Hz (ANSI S3.6, 1996).

B. Stimuli

The target speech material consisted of short every-day sentences, uttered by a male speaker (Versfeld *et al.*, 2000). The speech material comprised 39 lists of 13 sentences and was developed for a reliable measurement of speech intelligibility in noise. The speech was stored with a sampling rate of 44.1 kHz and 16 bits resolution.

Swedish and Dutch speech was used as interfering speech. The Swedish speech was developed by Hagerman (1982) and consisted of short sentences read by a female speaker. For the Dutch language, a corpus similar to that of Hagerman (1982) was developed. The Dutch sentences were uttered by a different female speaker. The Dutch speech was rescaled to obtain the same speaking rate as that of Hagerman's (1982) set, viz., 3.5 s per sentence. Scaling was done by use of the PSOLA method [Pitch-Synchronous Overlap Add, a method for manipulating the pitch and duration of an acoustic speech signal, Moulines and Charpentier, 1990]. Pseudorunning speech was obtained by concatenation of sentences in succession without pauses.

C. Procedure

Subjects were tested individually in a sound-insulated booth. The monaural speech-reception threshold was measured at the better ear for a fixed noise level of 65 dBA. Signals were played out via an Echo soundcard (Gina 24/96) on a PC at a sample frequency of 44.1 kHz and were fed through a TDT Microphone Amplifier (MA2) and a TDT Headphone Buffer (PA4) via THD 39P headphones. After the presentation of a sentence, the subject's task was to repeat the sentence he or she had just been presented. A sentence was scored as correct if all words in that sentence were repeated without any error. A list of 13 sentences, unknown to the subject, was used to estimate the level at which 50% of the sentences was reproduced without any error, the so-called speech reception threshold, or SRT. For a given condition, the first sentence of the list started far below the expected SRT. The sentence was repeated each time at a 4 dB higher level until the subject was able to reproduce it correctly. The 12 remaining sentences in that list were presented only once, following a simple up-down procedure with a step size of 2 dB. The SRT was estimated according to the procedure described by Plomp and Mimpen (1979), i.e., by taking the mean signal to noise ratio (SNR) of sentence 5 to 13 plus the

FIG. 1. Speech reception threshold (dB) as function of interfering speech type. Error bars denote the standard error of the mean.

estimated SNR that would have been used for the 14th sentence. With each sentence presentation, a random sample of the interfering speech was taken. It started 1200 ms before the start of the sentence and stopped at least 800 ms after the sentence.

In total four masking conditions were tested: Dutch masker forward, Dutch masker reversed, Swedish masker forward, and Swedish masker reversed. The experiment was partitioned into two blocks, a test and a retest. To avoid confounding of measurement condition order and sentence lists, the order of conditions and sentence lists was counter-balanced across subjects according to an 8 by 8 Latin Square method. In total, each subject received 8 lists of 13 sentences preceded by 3 practice lists.

III. RESULTS

Figure 1 shows the SRT values averaged across subjects and test-retest for each of the four conditions of the interfering speech. Error bars denote the standard error of the mean. A 4[condition] $\times$ 2[test/retest] $\times$ 8[subject] analysis of variance (ANOVA) was performed on the data set. Of the main effects, only differences in “condition” were significant ($F[3, 21]=24.6, p<0.001$). The SRT of the retest was on average 1.2 dB better than the test, but this difference was not significant ($F[1, 7]=4.6, p>0.05$). Also, differences between subjects were just not significant ($F[7, 4]=5.6, p>0.05$). None of the interactions were significant.

The difference in SRT between the two conditions with Swedish speech was significant (Tukey’s HSD test, $z=2.4, p=0.02$) and was on average 2.3 dB. Thus, by time reversing speech, the SRT increases (i.e., performance worsens) from -13.9 to -11.6 dB. The difference between the two conditions with Dutch interfering speech was also significant (Tukey’s HSD test, $z=4.28, p<0.001$). Here, time reversal of the masking speech causes the SRT to improve by 4.3 dB. The latter effect was apparent in all subjects, and the result is in agreement with that of Festen and Plomp (1990). Last, the

difference between the two conditions with reversed masking speech was also significant (Tukey’s HSD test, $z=3.6, p<0.001$).

IV. DISCUSSION

The most important finding of the present experiment is that time reversal of foreign, nonintelligible speech that is used as a masker, affects intelligibility and causes the SRT to increase by more than 2 dB. Most likely, reversal of the envelope of the speech signal results in an increased contribution of forward masking. The main conclusion to be drawn from this experiment is that reversal of a speech masker indeed introduces additional masking on top of energetic masking in the form of forward masking.

Time reversal of intelligible (Dutch) masking speech results in a decrease in the SRT of 4.3 dB. This difference is the result of the elimination of intelligibility of the masker on the one hand (enhancing speech intelligibility), and the introduction of additional forward masking on the other hand (decreasing speech intelligibility). If one assumes that the additional amount of forward masking is similar for the Swedish and Dutch masking speech, then in the present experiment the effect of informational masking due to intelligibility of the speech masker can be estimated to be $4.3 + 2.3 = 6.6$ dB.

If intelligibility of the speech masker were the sole factor determining the amount of informational masking, one would expect the SRT with Swedish reversed speech and Dutch reversed speech to be similar. In the present experiment this is not true; the difference is about 3.6 dB. Apparently, the Dutch reversed speech is a poorer masker (SRT = -15.2 dB) than the Swedish reversed speech (SRT = -11.6 dB). If Dutch interfering speech played in reverse sounds more similar to Dutch speech than Swedish speech played either forward or in reverse, one would expect the opposite. It is interesting to note that subjects told they could not distinguish between the Swedish reversed speech and the Dutch reversed speech. Also, they misjudged the reversed Dutch speech for Swedish speech and vice versa. Therefore, the effect probably is not very large. As mentioned in Sec. I, other factors than intelligibility contribute to informational masking. From the present results, it is difficult to say which of the many potential factors contributed most; perhaps differences in intonation, rhythm, mean pitch, modulation spectrum, or differences in the long-term speech spectrum. If the interfering native and non-native speech were uttered by the same person, with the same pronounced intonation, rhythm, and pitch, one would expect the SRTs in the reversed condition to be more alike. But even then, differences might still exist due to specific characteristics of the language.

The present experiment is a first attempt to separate energetic masking from informational masking. Future research with different native and non-native interfering speech, uttered by the same person may give more insight into the effects of forward masking and informational masking on speech intelligibility.

V. CONCLUSIONS

By using speech from different languages, played both forward or time-reversed, it is possible to disentangle the effects of informational masking caused by intelligibility of the interfering speech and energetic masking. However, time reversal of speech results in an increase in temporal (forward) masking. In the present experiment, this effect is about 2.3 dB. The release from informational masking by making intelligible speech unintelligible by reversing it in time is obscured by an increase in forward masking. In the present experiment, it is shown that informational masking might be as large as 6.6 dB if one assumes the same amount of additional forward masking with Dutch and Swedish reversed speech.

ACKNOWLEDGMENTS

Dr. B. Hagerman is acknowledged for providing us with the Swedish speech corpus. László Körösy is acknowledged for his help with the SRT computer program. We thank the associate editor, Dr. K.W. Grant, and the two anonymous reviewers for their constructive comments.

- ANSI (1996). "ANSI S3.6-1996, *American national standard methods for specification for audiometers*" (American National Standards Institute, New York).
- Bronkhorst, A. W. (2000). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acust. Acta Acust.* **86**, 117–128.
- Bronkhorst, A. W., and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.* **92**, 3132–3139.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Brungart, D. S., and Simpson, B. D. (2002). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J. Acoust. Soc. Am.* **112**, 664–676.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Carlyon, R. P. (1996). "Spread of excitation produced by maskers with damped and ramped envelopes," *J. Acoust. Soc. Am.* **99**, 3647–3655.
- Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (2001). "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment," *Audiology* **40**, 148–157.
- Drullman, R., and Bronkhorst, A. W. (2000). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," *J. Acoust. Soc. Am.* **107**, 2224–2235.
- Festen, J. M. (1987). "Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht), pp. 461–466.
- 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.
- Hagerman, B. (1982). "Sentences for testing speech intelligibility in noise," *Scand. Audiol.* **11**, 79–87.
- Irino, T., and Patterson, R. D. (1996). "Temporal asymmetry in the auditory system," *J. Acoust. Soc. Am.* **99**, 2316–2331.
- Moore, B. C. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
- Moulines, E., and Charpentier, F. (1990). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun.* **9**, 453–467.
- Peters, R. W., Moore, B. C., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**, 577–587.
- Plomp, R., and Mimpen, A. M. (1979). "Improving the reliability of testing the speech reception threshold for sentences," *Audiology* **18**, 43–52.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. R. Soc. London, Ser. B* **336**, 367–373.
- Schlauch, R. S., Ries, D. T., and DiGiovanni, J. J. (2001). "Duration discrimination and subjective duration for ramped and damped sounds," *J. Acoust. Soc. Am.* **109**, 2880–2887.
- Stecker, G. C., and Hafter, E. R. (2000). "An effect of temporal asymmetry on loudness," *J. Acoust. Soc. Am.* **107**, 3358–3368.
- Summers, V., and Molis, M. R. (2004). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," *J. Speech Lang. Hear. Res.* **47**, 245–256.
- 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.
- Versfeld, N. J., and Dreschler, W. A. (2002). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," *J. Acoust. Soc. Am.* **111**, 401–408.