On the perception of sinusoidally amplitude modulated signals and its relevance to listening in noise
Koopman, J.

Citation for published version (APA):
Koopman, J. (2004). On the perception of sinusoidally amplitude modulated signals and its relevance to listening in noise

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: http://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.
Chapter 7

The relevance of amplitude modulations to speech perception

In a silent environment, verbal communication presents few problems for normal hearing subjects or subjects with mild or moderate losses. Speech presented simultaneously with other speakers or subjected to reverberation, causes difficulties, especially for hearing-impaired subjects. The reason for the reduced performance of hearing-impaired subjects with regard to normal hearing subjects is primarily related to a reduced audibility. Since thresholds are elevated, speech is partly inaudible. However, at least part of the problem can be attributed to additional difficulties in segregating speech signals from background noise, due to an impaired temporal or spectral resolution (see Chapter 4).

This Chapter describes results, which illustrate the importance of the envelope to the intelligibility of speech. An important aspect is the ability of normal hearing subjects to improve speech intelligibility when modulated noise is present instead of continuous noise at the same signal-to-noise ratio. Hearing-impaired subjects obtain less or no benefit from the fluctuations in the masking noise. In the remainder of this thesis, the extent to which perception of speech by hearing-impaired subjects is limited by a reduced sensitivity to SAM is examined. Since speech is a highly
modulated signal, speech perception may be affected by MDI when speech is presented together with modulated maskers, such as fluctuating noise. Reduced masking release for speech (MRS) may be caused by excessive modulation masking.

### 7.1 Speech in noise

Both, noise and hearing-impairment, may reduce audibility. Intelligibility can be predicted by dividing the spectrum of speech into a number of bands. The signal-to-noise (S/N) ratio, or reduction in audibility due to hearing loss is then determined for each band. Weighing each band by a factor representing the importance of the band to overall intelligibility* and summing, results in the speech intelligibility index (SII). The SII ranges from 0 (not intelligible) to 1 (completely intelligible), giving the proportion of speech that is still available to the listener. Depending on the weights used, the SII can predict speech intelligibility for linear distortions such as a reduced audibility due to hearing-impairment or added noise. However, Pavlovic (1984) indicated that a model predicting intelligibility based on the audibility of the speech spectrum did not correctly predict intelligibility for subjects with moderate to severe hearing losses. This implies that factors, other than reduced audibility, may degrade the ability to understand speech. Noordhoek et al. (2000) showed that for subjects with hearing losses larger than 25 dB SL other parameters than audibility must limit the ability to understand speech. However, some subjects with larger losses may have ‘normal’ SII values. These subjects’ disability can be attributed to audibility alone.

#### 7.1.1 The Speech Reception Threshold (SRT)

A well accepted measure of speech intelligibility is given by the speech reception threshold (abbreviated as the SRT: Plomp and Mimp, 1979). The SRT is defined as the S/N ratio in decibels at which 50% of the sentences are repeated correctly. Performance strongly depends on the S/N ratio. Normal hearing subjects are unable to repeat sentences correctly for a broadband noise with the same long term average spectrum as the speech at a S/N ratio of approximately -8 dB, whereas all sentences

---

*Weighting factors and calculation routines are defined for octave bands, 1/3 octave bands, and critical bands.
are repeated correctly at a S/N ratio of -1 dB (Festen and Plomp, 1990; see open squares Figure 7.1). Hence, the range from completely intelligible to completely unintelligible is 7 dB. The SRT is about -5 dB. Hearing-impaired subjects generally perform worse than normal hearing subjects. This leads to higher SRT-values, ranging from -2 dB for subjects with mild or moderate losses to 10 dB for subjects with more severe losses. The range of S/N-ratios at which speech is completely intelligible and at which speech is not intelligible is comparable to what was reported for normal hearing subjects (see solid squares Figure 7.1). Although an elevation in SRT-score by 2-3 dB appears to be modest, it is sufficient to create a substantial loss of intelligibility in difficult situations.

7.1.2 Speech in fluctuating noise

Speech intelligibility is usually better for normal hearing subjects when the speech is presented in a fluctuating background noise (triangles) rather than a continuous background (squares) noise at the same S/N-ratio (Festen and Plomp, 1990; Howard-Jones and Rosen, 1993a; see Figure 7.1). When compared to normal hearing subjects (open symbols), hearing-impaired subjects (solid symbols) experience reduced intelligibility in continuous noise (squares). In addition, intelligibility does not increase when fluctuating noise is presented at the same S/N-ratio for hearing-impaired subjects.

Subjects benefited from the valleys in the masker leading to temporary improvements in S/N ratio. This improvement in speech intelligibility resulting from fluctuations in the background noise is referred to as masking release for speech (MRS) within the context of this thesis. The release of masking is largest for block-wave modulated noise at approximately 10 Hz (Miller and Licklider, 1950). In another study by Festen (1987), the SRTs for normal hearing and hearing-impaired subjects were obtained in a continuous background noise and a sinusoidally intensity modulated noise modulated by 4, 8, 16, and 32 Hz, having the same long term average spectrum as speech. For normal hearing subjects, the smallest masking release for speech (3 dB) is obtained for noise modulated by 4 Hz and the largest masking release for speech is found for modulation rates of 16 and 32 Hz (5.5 dB). These results presumably represent a trade off between the masking of complete words at low modulation rates and insufficient temporal resolution at higher modulation
Figure 7.1: Average discrimination curves for sentences presented in steady state noise (squares) and fluctuating noise (triangles), for normal hearing (open symbols) and hearing-impaired listeners (solid symbols) (based on 140 responses) redrawn from Festen and Plomp (1990)

rates. However, for hearing-impaired subjects, the SRT is 3.2 dB higher than the SRT in continuous noise. This is attributed to the reduced temporal resolution of the hearing-impaired subjects relative to normal hearing subjects. Trine (1995) measured the effect of Gaussian noise, interrupted at rates of 2, 4, 8, 16, 32, 64 and 128 Hz, on intelligibility for normal-hearing subjects. Subjects were asked to adjust the level of the noise until they thought that 50% of the speech was intelligible. Masking release for speech was approximately 22 dB for modulation rates below 16 Hz and decreased monotonically to 0 dB for a rate of 128 Hz. Gustafsson and Arlinger (1994) determined that, by using SAM noise as a masker, higher modulation (> 10 Hz) rates produced less masking release than lower modulation rates and modulation depths of -6 dB \( (20 \log_{10}(m)) \) resulted in a smaller release of masking than modulation depths of -12 dB or 0 dB (100%).
The most obvious example of a fluctuating background noise is given by a competing speaker\(^1\). Hearing-impaired subjects often report great difficulties understanding speech with one interfering talker. The masking release for speech found for normal hearing subjects (difference in squares and triangles) is approximately 7 dB. For hearing-impaired subjects, SRTs for a fluctuating and a continuous background noise usually are comparable. Hence, the difference in the SRT using a fluctuating background noise can be over 12 dB.

The benefit of the fluctuations in noise can only partly be attributed to a reduced audibility. Bacon et al. (1998) measured speech intelligibility for normal hearing subjects in a fluctuating noise to which a continuous background noise was added to simulate the reduced audibility of hearing-impaired subjects. The masking release for speech reduced, but fluctuations in the background noise were still beneficial to normal hearing subjects. This implies that the reduced intelligibility in a fluctuating noise in hearing-impaired listeners is at least partly due to supra-threshold deficits.

7.2 Factors limiting the intelligibility in noise

The supra-threshold deficits limiting speech intelligibility can be attributed to the altered spectral and temporal resolution of hearing-impaired subjects (Noordhoek et al., 2001). The effect of hearing-impairment on these measures are discussed in Chapter 4 of this thesis. Two methods can be used to study the effects of these psychoacoustical factors on speech intelligibility:

1. **Correlational studies.** The correlation between the performance on an auditory task and speech intelligibility is determined.

2. **Disturbing the speech signal.** Disturbing the speech signal may provide information on the cues that are important to speech intelligibility. This information can be obtained by:

\[^1\]The fluctuating noise used to obtain the data in Figure 7.1, is created by filtering concatenated sentences at 1 kHz and replacing the fine structure of the low and high-pass segments with a continuous noise leaving the temporal envelope intact.
(a) determining the effect of specific distortions, such as reverberation, on intelligibility by leaving the speech signal intact.\footnote{Applying distortions to a signal will generally alter the signal within certain limits. For instance, applying a gain to certain frequency components will alter the temporal structure and phase relations.}

(b) studying the effect of reducing information within the speech signal on intelligibility by, for instance, filtering.

### 7.2.1 Correlational studies

A variety of studies have examined the relationship between the intelligibility of speech presented in noise and a diversity of spectral and temporal psychoacoustical parameters.

Festen and Plomp (1983) indicated that measures of spectral resolution, such as critical ratio and the low frequency edge of the estimated PTC, are clearly related to the intelligibility of speech in noise. These findings have been confirmed using parameters such as the critical ratio (Dreschler, 1983), the width of the psychoacoustical tuning curve (Horst, 1987) and just noticeable differences in frequency (Glasberg and Moore, 1989). Festen and Plomp (1983) determined speech intelligibility related to measures of temporal resolution, such as forward masking, backward masking and click thresholds in noise. Relatively weak correlations were reported. This has been confirmed by other studies and for other measures of temporal resolution, such as the sensitivity to SAM (Takahashi and Bacon, 1992) and forward and backward masking (Dreschler, 1983). However, results for gap-detection are inconclusive. Modest correlation are reported for speech in noise (Dreschler, 1983) or relative to the STI in a reverberant room (Dreschler and Lecuy, 1990). However, strong effects have been reported by Glasberg and Moore (1989).

Speech intelligibility with correction of audibility (SII) was measured with regard to auditory performance for tasks measuring temporal or spectral resolution by Noordhoek et al. (2001). Positive significant correlation coefficients were reported, suggesting that supra-threshold deficits indeed limit speech intelligibility. The components of a PCA on the psychoacoustical measures indicated that three components, temporal resolution (shape discrimination, forward and backward masking), spectral discrimination (frequency discrimination), and spectral resolution
7.2. Factors limiting the intelligibility in noise

(upward and downward spread of masking) contributed significantly (p-value < 0.05) to speech intelligibility in noise.

**CMR related to the intelligibility of speech**

Co-modulation Masking Release (CMR) refers to the increased ability to detect a signal in noise using information on the envelope of the masker outside the filter tuned to the target. Howard-Jones and Rosen (1993b) measured the effect of 'uncomodulated glimpsing' on masking release using a 'checkerboard' noise. The masking release for a fully co-modulated background noise is approximately 23 dB and decreases when multiple bands are modulated in anti-phase and is absent for 8 or more spectral bands. When only two bands are used, the 'checkerboard' noise results in a higher masking release than when the noise contains one modulated channel and one continuous channel. Apparently, listeners benefit from 'uncomodulated glimpsing' by combining speech cues from different frequency bands at different times in the signal. In section 4.4.2, literature has been discussed that indicated that CMR vanished when masker envelopes modulated 180° out of phase were used. In addition, using speech as a target signal introduces two differences with traditional CMR (Grose and Hall, 1992: see section 4.4):

1. the target (speech) is a complex broadband signal rather than a spectrally discrete signal (tonal).

2. the tasks (speech intelligibility) involves supra-threshold recognition rather than detection.

Festen (1987) used noise divided into two 100% SIM modulated bands running in and out of phase. No clear benefit for co-modulation at 4 Hz was reported and a relatively small benefit of approximately 1 dB at the highest modulation rate (32 Hz). Grose and Hall (1992) reported a clear effect of CMR in the ability to detect speech, whereas no clear effect of CMR on intelligibility was found. Festen (1993) measured speech intelligibility by masking speech with speech divided into multiple

---

5) broadband noise was split into multiple bands (2, 4, 8, or 16) of equal power and modulated either in phase or out of phase at a 10 Hz rate. A spectogram of this type of signal results in an image that looks like a checkerboard

6) Festen used sinusoidally intensity modulated noise, in order to approach the effect of an interfering speaker more realistically.
bands. The effect of CMR, determined by the difference in intelligibility for speech presented in a masking voice with and without time delay between 1/3 octave bands, was small. Again, the phases of the envelopes are crucial to obtain a fair amount of CMR, and the weak relationship between CMR and masking release for speech may not indicate that the relationship does not exist. Kwon (2002) measured the recognition of band-filtered consonants masked by multiple bands and observed a small effect (3.5%) for CMR.

CMR can be obtained for signal detection tasks (Hall and Grose, 1988) or the detection of speech (Grose and Hall, 1992). However, the literature is inconclusive as to whether CMR occurs for supra-threshold recognition. For tasks such as gap detection, conditions benefitting CMR increase the sensitivity to the gap (Hall and Grose, 1992). In contrast, the sensitivity to other supra-threshold recognition tasks such as speech intelligibility (Grose and Hall, 1992; Festen, 1993; Kwon, 2002) and amplitude modulation discrimination (Hall and Grose, 1995) did not increase using conditions that enhance CMR.

**MDI related to the intelligibility of speech**

A limited number of studies have focused on the intelligibility of speech in relation to measures as MDI. This is unsurprising since the intelligibility of speech improves, while the sensitivity to SAM is reduced when modulators maskers are used. However, MDI may reduce the masking release for speech. Takahashi and Bacon (1992) reported weak correlation coefficients between modulation detection and masking release for speech, expressed in percentage correct at a S/N ratio of 0 dB. Kwon and Turner (2001) measured the intelligibility of filtered consonants presented in fluctuating masker bands at adjacent frequencies. When the speech and masker were widely separated in frequency, the effect of MDI was dominant.

**MDI and masking release for speech: results from Chapter 5**

Binaural speech intelligibility for the subjects, who participated in the MDI experiments described in Chapter 5, was determined using a standard SRT-test. Figure 7.2 illustrates the results of the SRT in a fluctuating noise (ordinate) as a function of the SRT in a continuous background noise (abscissa). The diagonal line indicates where the points would fall if identical results were obtained for speech in
7.2. Factors limiting the intelligibility in noise

Figure 7.2: Scatter plot of individual speech intelligibility performances in a continuous (abscissa) and a fluctuating background noise (ordinate). The diagonal line indicates the points for which the SRT in a fluctuating noise equals the SRT in a continuous noise (masking release for speech = 0).

a continuous and in a fluctuating noise. Hence, masking release for speech is given by the deviation from the diagonal line and is larger for normal hearing than for hearing-impaired subjects.

Since the number of subjects is small, rank-order correlation coefficients were determined (Spearman’s $\rho$) in order to determine whether masking release for speech and MDI are linked. The results suggest that there is no relationship between MDI and masking release for speech (see Table 7.1). The only exception is the reduced sensitivity to SAM with regard to the conditions without maskers ($MDI_{PA}$) using a reference depth of 0.30 ($p$-value < 0.05). Figure 7.3 gives the scatter plot between these two parameters. The figure indicates that more interference coincides with a larger MRS, which is in contradiction with the stated hypothesis. Apparently, the reduction in MRS is not related to an excessive amount of modulation masking.

However, determining the Spearman rank correlation coefficients for normal hearing subjects and hearing-impaired subjects separately, indicates that both groups have a negative non-significant correlation coefficient for reference depth 0.18.
Chapter 7. The relevance of AM to speech perception

| MDI_{UNM}(0.18) | 0.07 |
| MDI_{UNM}(0.30) | 0.16 |
| MDI_{PA}(0.18)  | 0.56 |
| MDI_{PA}(0.30)  | 0.62*|
| MDI_{UNM}       | -0.14|
| MDI_{PA}        | 0.24|

Table 7.1: Rank correlation coefficients between the masking release for speech (MRS) and the amount of interference (MDI). The masking release for speech was determined using the difference in the SRT for speech in a fluctuating and a continuous background noise. The data between brackets give the reference depth. MDI_{UNM} gives the reduced sensitivity to SAM by adding modulated maskers instead of non-modulated maskers. MDI_{PA} is determined relative to the condition without maskers. Asterisks denote the level of significance, p-value<0.05 *; p-value<0.01 **.

Figure 7.3: Scatter plot of the release of masking for speech (ordinate) against the interference caused by the modulated maskers at a reference depth of 0.30 (abscissa). The solid line shows the regression line between the data.
This indicates that there is a tendency for masking release for speech to decrease as MDI increases. However, for reference depth 0.30, both groups exhibit positive correlation coefficients. All correlation coefficients indicate non-significant relations.

<table>
<thead>
<tr>
<th></th>
<th>MDI (0.18)</th>
<th>MDI (0.30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>( \rho_{NH} )</td>
<td>-0.54</td>
<td>0.18</td>
</tr>
<tr>
<td>( \rho_{HI} )</td>
<td>-0.15</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 7.2: Rank correlation coefficients between the masking release of speech and the amount of interference.

More on the use of fluctuations in the masking signal

The SRT-test is a valuable method for determining the subjects’ intelligibility. A small disadvantage of the method is that it does not provide a good insight into the relationship between intelligibility scores (percentage correct) and the S/N ratio used. These curves are given by Festen and Plomp (1990). The performance in percentage is given for normal hearing and hearing-impaired subjects as a function of the S/N ratio of speech presented in a continuous and fluctuating noise (see also Figure 7.1). The results indicate that normal hearing subjects experience improved intelligibility when the masking signal contains gaps, and the benefit of these gaps increases as percentage correct-scores decrease (the curve for speech in fluctuating noise is more shallow than the curve for continuous noise). The gaps in the noise do not improve speech intelligibility for hearing-impaired subjects. In addition, the slopes for the continuous noise and the fluctuating noise are comparable to each other.

In order to assess the ability to use the dips of the fluctuations in a background noise ('glimpsing'), matching experiments have been conducted (not described in this thesis) in which subjects are asked to adjust the level of the speech until speech was:

1. just inaudible \([net niet hoorbaar]\)
2. audible \([hoorbaar]\)
3. well audible, not intelligible  
   \[\text{goed hoorbaar, niet verstaanbaar}\]
4. just intelligible  
   \[\text{net verstaanbaar}\]
5. very intelligible  
   \[\text{goed verstaanbaar}\]

Figure 7.4: Results of matching experiments carried out for normal hearing subjects (9 subjects; left panel) and hearing-impaired subjects (3 subjects; right panel) carried out in a fluctuating noise (triangles) and a continuous background noise (circles)

Experiments were carried out in a continuous and in a fluctuating speech-shaped noise in order to determine the masking release for speech. Figure 7.4 shows the mean of the differences for the subjects’ S/N-ratio and their SRT for that noise. Hence, the data have been corrected for masking release for speech as determined by the SRT measurements and the difference between the two curves represents the additional benefit of fluctuations. The results clearly indicate that the fluctuating background noise provides additional masking release for the detection of speech for normal and hearing-impaired subjects (diverging curves). As aforementioned, CMR effects have been reported for speech detection, but not for speech intelligibility (Grose and Hall, 1992). However, the monotonic converging curves in Figure 7.4 illustrate that the difference between the intelligibility and the ability to detect speech does not show a sudden increase in masking release for speech.

Given the importance of the coherence in the envelope to CMR, it seems plausible to expect a similar importance of coherence for the target signal. Although
the envelope of speech is roughly coherent over frequency bands. Figure 7.5 clearly indicates that, for a randomly chosen sentence, different bands within speech are not completely coherent. There are large contrasts in the amount of energy in the periods indicated by the vertical lines in the different frequency bands. CMR is known to decrease as phase differences for across-band information (the masker) increase. Therefore it is questionable whether CMR can account for the improvement in the intelligibility of speech due to the fluctuating behavior of the masker.

![Figure 7.5: A Dutch sentence filtered into octave bands of 0.5, 1, 2, or 4 kHz. Vertical lines are used to indicate periods in which there is little coherence between bands.](image)

### 7.2.2 Disturbing the speech signal

**Interference of the intact speech signal**

The relationship between speech intelligibility and a distorted auditory coding can be determined using the distortion sensitivity approach (Plomp, 1986). Removing cues that cannot be used due to a distorted auditory coding will not affect speech intelligibility for hearing-impaired subjects. Measuring intelligibility as a function of the distortion applied to the signal will reduce the difference in performance for
normal hearing and hearing-impaired subjects. However, when the difference in intelligibility scores for normal hearing and hearing-impaired subjects is constant, independently of the applied distortion, hearing-impaired subjects are as sensitive to the distortion as normal hearing subjects. This indicates that the distortion is not related to the deficits limiting intelligibility of the individual hearing-impaired subject.

Duquesnoy and Plomp (1980) determined the effect of reverberation on speech intelligibility as a function of the reverberation time. Their results indicate that hearing-impaired subjects are as sensitive as normal hearing subjects to reverberation. As a tool to describe auditory processing van Schijndel (2000) used wavelet coding\(^1\). A distortion sensitivity model was used to determine the effect of temporal, spectral and level distortions on speech intelligibility. Results indicate that hearing-impaired subjects are as sensitive as normal hearing subjects to a distorted coding of temporal or level information. This suggests that hearing-impaired subjects do not experience additional problems as a result of an altered coding strategy with regard to temporal and level distortion. Applying spectral distortion, clearly indicated that speech perception for hearing-impaired subjects was reduced less than for normal hearing subjects. This indicates that the problems hearing-impaired subjects experience in the intelligibility of speech are mainly caused by sub-optimal spectral coding properties.

To investigate the relative importance of envelope and fine structure, Smith et al. (2002) generated signals that have the envelope of one sound and replaced the original fine structure by the fine structure of the other. The envelope appeared to be most important to speech intelligibility, whereas the fine structure is most important for pitch perception (melody recognition) and sound localization.

**Reducing information in the speech signal (spectral)**

In section 7.1 the effect of adding noise has been discussed in terms of the SII. The SII assumes that overall intelligibility is determined by the S/N ratio within each

\(^1\)Wavelet coding is a tool to describe the temporal-spectral behavior of the peripheral part of the ear. To some extent, it can be compared to the short-time Fourier analysis. However, the spectral resolution for wavelet coding decreases with frequency (constant on a log-frequency scale). This is more similar to the coding strategy of the auditory system. Experiments also suggested that the duration of the time-frequency window decreases with frequency.
7.2. Factors limiting the intelligibility in noise

band. with each band having its own importance weighting. The band importance functions were introduced to account for the effect that some bands contribute more to overall intelligibility than other bands (Pavlovic. 1987). Intelligibility is determined for each band independently, with 1/3 octave bands between 800 Hz and 4 kHz having the highest weights. These frequencies represent 80% of the information in speech. Lippman (1996) indicated that reasonable intelligibility was obtained for CVC-words for frequencies below 800 Hz (44%). However, spectral information above 8 kHz, still offers a significant increase in CVC-score of approximately 30% in intelligibility. Hence, speech is a multi-channel process for which different channels fuse and lead to a single percept. Such interaction is also reported for across-channel processes such as CMR and MDI.

Reducing information in the speech signal (temporal)

Shannon et al. (1995) studied the extent to which fine structure information is essential for speech intelligibility. The original fine structure of speech was replaced by a white noise, while maintaining the envelope information for each band. The number of bands ranged from one to four**. The intelligibility of consonants, vowels, and words is above 80% using three bands. Using four bands increased performance to about 90%, which indicates that nearly perfect intelligibility can be obtained using the temporal structure of just four bands. Replacing the carrier signal of speech by pure tones modulated by the envelope of the band centered at the signal frequency still results in a fair intelligibility (Dorman et al., 1997).

Disturbance of the temporal structure whilst keeping the fine structure intact has been studied by Drullman et al. (1994a,b). The speech signal is divided into spectral frequency bands. The envelope of each band was low-pass or high-pass filtered. Filtering out modulation frequencies above 16 Hz barely affected the intelligibility of consonants, vowels and sentences. Filtering out modulation frequencies below 16 Hz gradually reduced speech intelligibility (Drullman et al., 1994b). Filtering out modulation frequencies below 4 Hz also barely affected the intelligibility of consonants, vowels and sentences. Filtering out frequencies above 4 Hz reduced the intelligibility of speech gradually (Drullman et al., 1994a). When

** All conditions were low-pass filtered at 4 kHz. Filter cutoff frequencies were 1500 Hz for the two-band processor, 800 and 1500 Hz for the three-band processor, and 800, 1500, and 2500 Hz for the four-band processor.
the two studies are compared, they indicate that the effect of low-pass and high-pass filtering of the envelope is approximately the same at 8-10 Hz. A combination of previous studies has been studied by van der Horst et al. (1999). The consonants of VCV syllables were split into bands of which the original fine structure was replaced by a speech-shaped noise. For each band, modulation frequencies were filtered out around 8, 12 and 16 Hz. A large reduction was found for relatively wide notches (>10 Hz) around 8 Hz, whereas other conditions result in a marginally decreased performance.

The importance of temporal troughs was studied by Drullman (1995). The troughs were flattened, simulating the effect of added noise. The SRT was 5 to 6 dB better for manipulated speech than for speech in noise, which indicates that the added noise may confuse the listener, for instance by the inherent fluctuations of the added noise. Replacing the fine structure of speech by the fine structure of noise, leaving the temporal speech envelope intact indicates that speech intelligibility is barely affected by manipulations of the fine structure. However, maintaining the fine structure and randomizing the envelope causes intelligibility to drop to 17%.

Another aspect of the temporal structure of speech was investigated by Greenberg et al. (1998). Speech is divided into one-third octave bands and intelligibility was measured using these bands alone or in combination with each other. Intelligibility was high, even with only three bands. However, desynchronizing these bands by more than 25 ms, severely degraded speech intelligibility. Increasing the number of bands to 19, made listeners relatively insensitive to temporal asynchrony (Arai and Greenberg, 1998).

### 7.3 Implications

This chapter has showed that:

1. Masking release for speech is substantial for normal hearing subjects and absent for hearing-impaired subjects.

2. The temporal envelope of spectral bands in speech is an important factor in the intelligibility of speech.

3. The intelligibility of sentences is high when the temporal envelope of speech
7.3. Implications

consists of four bands or more.

4. speech is a very robust signal and considerable temporal distortion is needed to reduce speech intelligibility.

Hence, speech is a highly redundant signal, in which perceptual trade-offs occur. Information across auditory filters provides additional cues, which improve speech intelligibility. Given the importance of the envelope to the intelligibility of speech, the results described in section 7.2.1 are surprising.

The lack of evidence supporting a relationship between MDI and MRS may be due to the difference in gating. In the MDI-experiment, target and masker were gated synchronously, whereas speech intelligibility was assessed using asynchronous gating of speech and masker. Previous studies have indicated that MDI is largest when the probe and masker carriers are gated synchronously, with the amount of interference falling as gating asynchrony is increased (Hall and Grose, 1991; Mendoza et al., 1995b). However, even for continuous gating, a fair amount of MDI was measured. This difference in gating may contribute to the low correlation coefficients. The results described in Chapter 5 indicated large differences between normal hearing and hearing-impaired subjects in the tendency to adjust the signal with a larger modulation depth than the reference depth. This tendency may reflect a reduction in sensitivity to SAM due to pre-exposure to SAM. Speech intelligibility may be affected and the effect may be considered as ‘true’ MDI, indicating that a reduced sensitivity to SAM also results from factors other than perceptual grouping.