SUMMARY

Motivated by the observation that most noise reduction algorithms fail to improve the intelligibility of speech in background noise, the aim of the present thesis was to identify the crucial factor – in signal-analysis terms – for the intelligibility of noise-corrupted speech. By definition, such a factor must be independent of any type of signal processing applied to the noise-corrupted speech signal. Recent publications have shown that this is not the case for common measures like signal-to-noise ratio (S/N ratio) and speech transmission index (STI). It has been observed that S/N ratios and STI values of noise-corrupted speech can improve as a result of signal processing, while the intelligibility of the processed noisy speech remains equally poor. Until this paradox is unravelled there remains a gap in the theory of noisy speech perception, and it remains unclear what principle noise-reduction algorithms should aim for in terms of signal restoration. To understand this paradox is the driving force behind the current thesis, in which the physical and psychophysical effects of noise on speech and its perception were studied in detail. The general idea is that when noise is added to speech, the speech waveform is altered in various ways. Each alteration has a specific degrading effect on intelligibility, some more then others. Once the most degrading effect is identified, we can attempt to improve the intelligibility of noise-corrupted speech in general by counteracting particularly this most prominent effect of noise.

After a general introduction in Chapter I, a specific type of signal analysis is introduced in Chapter II, designed for identifying and isolating the different speech-waveform alterations when noise is added. As it was intended to relate signal processing to auditory processing, the applied signal analysis was matched to human auditory analysis with respect to the applied spectro-temporal resolution. It is known that the spectral resolution of the auditory system is roughly constant on a logarithmic frequency scale. Wavelet analysis is well suited for a logarithmic frequency scale, and was used in this study to capture the instantaneous amplitude and phase of 1/3-octave band signals, in a way roughly corresponding to the auditory bandwidth. After subjecting speech-plus-noise to a wavelet analysis, three different effects of noise on the wavelet representation were identified: two effects on the wavelet level (one systematic and one stochastic) and one effect on the wavelet phase. Using signal processing, each of these three effects could be included or excluded in speech stimuli. The resulting eight conditions were subjected to intelligibility tests (CVC word scores), which indicated that the systematic level effect (i.e. increase of each speech wavelet with the mean noise intensity) has the most detrimental effect on intelligibility, causing a drop in word score from 83% (no noise) to 63%. The phase effect appeared less degrading (76%), while stochastic perturbation of the speech envelope (86%) did not seem to have any negative effect by itself, only in combination with a second effect did it affect scores. In general, adding a second effect reduced the intelligibility further. Up to this point, the results seemed in good agreement with models like the MTF-STI. In this model, the effect of noise on speech intelligibility is understood as a reduction of speech modulations essential for intelligibility, expressed in the reduced peak/valley ratio in the fluctuating speech envelope caused by noise filling up the speech valleys. On average, the valleys are raised by an amount equal to the mean noise intensity, which corresponds exactly to the 'first noise effect' found in our experiment.

So, increasing the intelligibility of noise-corrupted speech seems basically a matter of neutralizing the first noise effect. In signal-technical terms, this operation is identical to applying all noise effects excluding the systematic level effect, which was one of the eight measured conditions. Results indicated that intelligibility indeed increased, from 29% ('full noise effect') to 74% after applying the all noise effects but the main one, confirming the significance of neutralizing the first noise effect.

For practical applications, however, the applied operation is strongly limited by the fact that access to the original, uncorrupted, speech signal is required. An alternative operation, in which an estimate of the mean noise intensity was directly subtracted from the intensity of the speech-plus-noise mixture – a basic form of 'spectral subtraction' – did improve STI

values, indicating successful neutralization of the first noise effect, but failed to improve intelligibility.

To try and understand the completely different perceptual effects for both types of noise reduction (with and without *a priori* knowledge), the output signals of both operations were compared, showing completely different details in their effects on the signal. The first operation (using original speech) restored the global contours of the speech envelope, and thus effectively reduced the noise fluctuations relative to those contours. The second operation (spectral subtraction) compensated for the overall level increase caused by the noise, but did not effectively change the distribution of the speech and noise fluctuations. As a result, the strength of noise modulations was strongly reduced in the first case, but remained essentially unchanged in the second case, suggesting that the intelligibility of speech in noise may not only depend on the strength of speech modulations, but also on the strength of noise modulations. It was suggested that intelligibility might depend on the ratio between speech modulations and noise modulation.

To verify this idea, noise-corrupted speech was subjected to various types of processing (spectral subtraction, deterministic modulation reduction, amplitude compression and expansion), and the ratios between speech and noise modulations (the "modulation ratio") of the processed signals were compared with intelligibility scores obtained from listening experiments (described in Chapter III). As the modulation ratio cannot be determined directly from the noisy speech itself, a special probe signal was designed for this purpose which, mixed with noise, was subjected to the various signal processing types. For each processing type, modulation ratios were determined for this probe signal, and compared to intelligibility scores of noisy speech that was processed in the same way. The results of this study corroborated the relation between the modulation ratio and intelligibility, and motivated the introduction of a new concept: the signal-to-noise ratio in the modulation domain, or $(S/N)_{mod}$.

Increasing the intelligibility of noise-corrupted speech appears to be basically a matter of increasing the $(S/N)_{mod}$, i.e. increasing the strength of speech modulations *re* noise modulations. To further investigate this concept, the $(S/N)_{mod}$ of noisy signals was artificially manipulated by means of specific signal processing, and again compared to intelligibility scores obtained from listening experiments (Chapter IV). This time, the artificial probe (used to estimate $(S/N)_{mod}$) was adjusted to match the intensity distribution of actual speech, allowing $(S/N)_{mod}$ values to be converted into the expected SRTs. In the first part of the Chapter, the imposed variations in $(S/N)_{mod}$ were controlled with *a priori* knowledge about the original (uncorrupted) signal. As this information is only available in the laboratory, an alternative approach was tested in the second part of the Chapter, in which $(S/N)_{mod}$s were manipulated without this type of information. It was shown that the variations imposed on the $(S/N)_{mod}$s of the signals were essentially followed by corresponding variations in intelligibility, substantiated by correlation coefficients of typically 0.8. It is argued that the applied type of signal processing may be useful for future practical applications, as the results indicated that (without *a priori* knowledge) intelligibility improvements of typically +2 dB could be reached for hearing-impaired persons. For practical applications of this type of processing, it is important to know whether improvements can also be shown for more realistic noise types, such as, for example, babble noise. To obtain some clarification, an experiment was performed (Chapter V) using realistic babble noise obtained from a live recording of about fifty talking persons at a lively reception. Nine normally-hearing and sixteen hearing-impaired persons participated in the experiment, which was performed in essentially the same way as the previous experiments. The results indicated that the processing still improved intelligibility for normally-hearing and hearing-impaired persons, on average by 0.6 and 1.6 dB, respectively. It was also found that the predictions based on the $(S/N)_{mod}$ did not apply well for small numbers of interfering talkers, as the present concept does not account for the release of masking observed for such strongly fluctuating maskers. This limitation of the applicability of the $(S/N)_{mod}$ concept requires further investigations.