Speech Communication 12 (1993) 261–266
North-Holland

Intelligibility improvement of noisy speech for people with cochlear implants

H.G. Hirsch

Institut für Nachrichtengeräte und Datenverarbeitung, Aachen University of Technology, Templergraben 55, 52056 Aachen, Germany

Received 18 March 1993

Abstract. With a cochlear implant, deaf people are able to understand speech under good listening conditions. Problems occur in adverse conditions, e.g. in reverberant and/or noisy environments. Experiments were carried out to improve the intelligibility of noisy speech using two different single-channel noise suppression techniques. This was realized by preprocessing, i.e. by applying the resynthesized speech signal to the cochlear implant system. Intelligibility tests were carried out in cooperation with the medical department.

Zusammenfassung. Ertaubte Personen sind mit Hilfe eines Cochlea Implantats in der Lage, Sprache in einer ruhigen Umgebung zu verstehen. Es treten Probleme in gestörten Hörsituationen auf, z.B. in halliger und/oder störschallerfüllter Umgebung. Einige Untersuchungen zur Verbesserung der Verständlichkeit gestörter Sprache wurden angestellt. Dazu wurden zwei unterschiedliche, einkanalige Störunterdrückungsverfahren verwendet. Realisiert wurde dies in Form einer Vorverarbeitung, wobei die resynthetisierte Sprache dem Cochlea Implantat System zugeführt wird. Verständlichkeitstests wurden in Zusammenarbeit mit dem medizinischen Fachbereich durchgeführt.

Résumé. Avec une cochlée artificielle, des personnes mal entendantes sont capables de comprendre la parole lorsque les conditions d'écoute sont bonnes. Toutefois, en présence de bruit ou de réverbération la compréhension pose des problèmes. Des tests ont été menés pour améliorer l'intelligibilité de la parole bruitée en utilisant deux techniques différentes de suppression de bruit à un seul canal. Ceci fut réalisé par prétraitement, c'est a dire en réinjectant a la cochlée artificielle le signal de parole résynthétisé. Des tests d'intelligibilité ont été menés en collaboration avec un service médical.

Keywords. Speech enhancement; electronic hearing aid.

1. Introduction

Electronic hearing aids have been developed during the last twenty years for deaf people with a postlinguistic hearing loss. These are called cochlear implant systems. The main component is an array of electrodes which is implanted in or nearby the inner ear. The electrodes are stimulated by a speech processor which carries out some speech analysis and coding algorithm. The detailed structure will be shown in the following section.

Deaf people can understand continuous speech again using only a cochlear implant or using it in combination with lip reading. Because of the limited number of electrodes and because of the simple speech analysis technique, these people have considerable problems to understand speech in adverse conditions, especially in the cases of reverberation and/or additive noise. The degradation of speech intelligibility is much more severe than for normal binaural hearing people. Therefore, more sophisticated speech preprocessing techniques should be applied.

0167-6393/93/$06.00 © 1993 – Elsevier Science Publishers B.V. All rights reserved

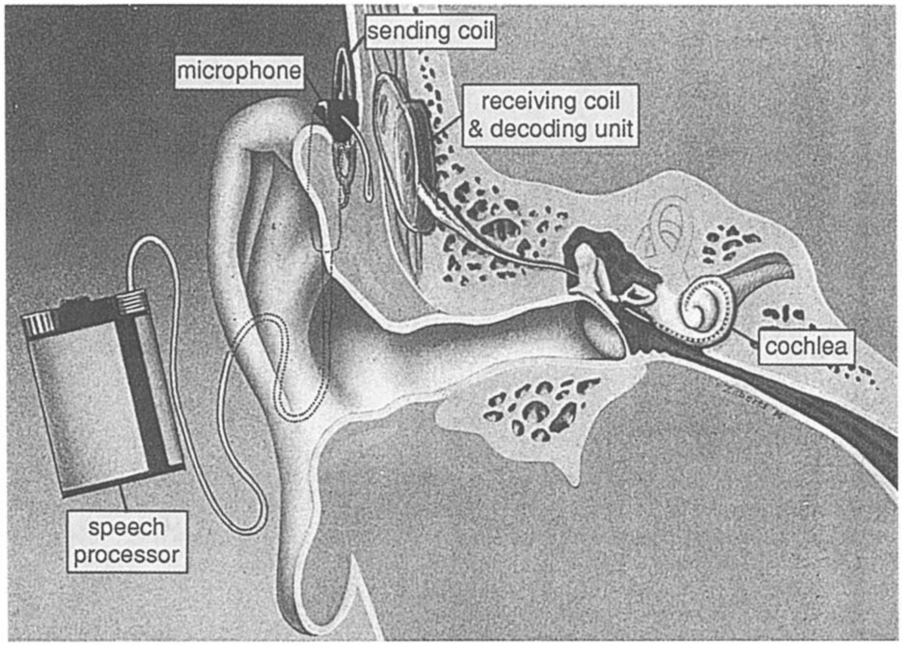

Fig. 1. Structure of the cochlear implant system. (Reproduced with permission form Cochlear AG, Switzerland, Artwork by Mr. E. Zilbert.)

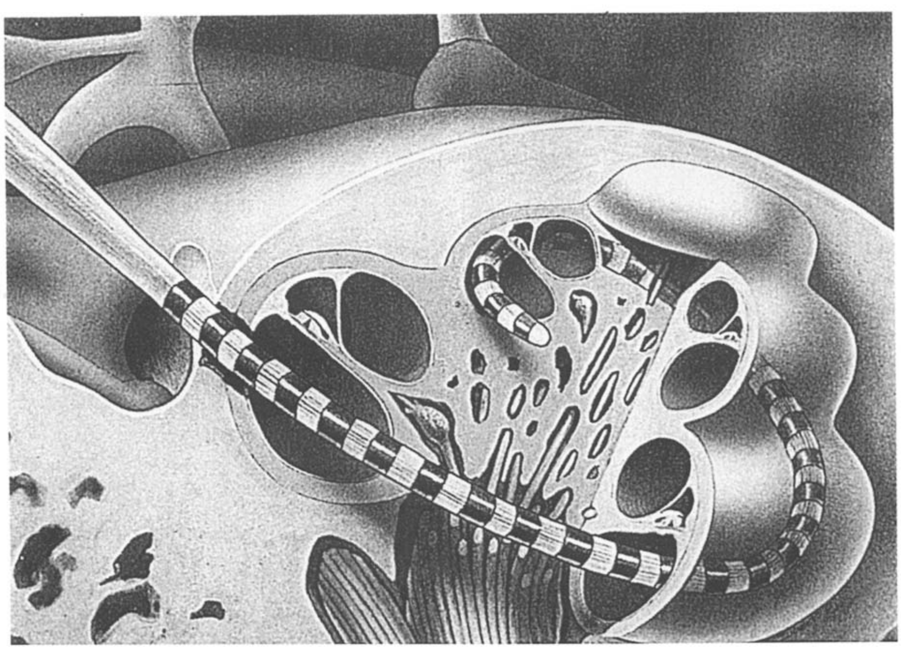

Fig. 2. Array of electrodes implanted in the cochlea. (Reproduced with permission from Cochlear AG, Switzerland, Artwork by Mr. E. Zilbert.)

In a first approach additive, nearly stationary noise was examined. We used a noise suppression algorithm for preprocessing. Intelligibility tests were carried out with a few people.

2. Cochlear implant

Different concepts were developed and examined for the realization of cochlear implants (Clark et al., 1990). They differ in the number of electrodes, the place of implantation and excitation (intra-/extracochlear), the kind of excitation (analogue/pulses) and the extracted acoustic features. This study was done in cooperation with the medical department in Aachen. An intracochlear implant with 22 electrodes is used called "Nucleus 22", which had been developed at the University of Melbourne (Clark et al., 1990; Cochlear, 1989). The complete system is shown in Figure 1.

Acoustic signals are recorded with a directional microphone placed directly above the ear. The signal is transmitted to a speech processing unit for analysis and coding. The extracted acoustic parameters are coded as pulse series of a digital signal with a frequency of 2.5 MHz. This signal is transmitted to a sender coil which is placed behind and above the ear. The coil is fixed by a magnet of the implanted receiver-stimulator unit. This unit consists of a receiving antenna, a decoder circuit and a stimulation unit which generates the electrical pulses for the electrodes.

The array of electrodes is implanted in the inner ear as shown in Figure 2. The electrical pulses lead to a stimulation of auditory nerve fibres. Selecting e.g. a specific pair of neighbouring electrodes, nerve fibers in the region of the electrodes position are stimulated corresponding to the well known frequency-to-place transformation in the inner ear.

The speech processing unit analyses the speech signal. Some spectral parameters and the pitch period are estimated. For voiced sounds two formant frequencies, the corresponding energy and the energy in the regions of 2000–2800 Hz and 2800–4000 Hz are calculated. The corresponding electrodes are stimulated with the rate of the estimated pitch period. For unvoiced sounds a random pitch and the energy above 4000 Hz instead of the first formant are used.

The speech processing unit offers an internal noise suppression mechanism. This is described as an adjustment of the threshold stimulation level dependent on the constant level of background noise (Cochlear, 1989). It can be imagined as a simple spectral subtraction in only two broad bands which are used for the estimation of the formant frequencies.

3. Noise suppression with a single microphone

Many algorithms exist for the preprocessing of speech to reduce the influence of noise and/or reverberation, e.g. (Lim, 1983; ESCA, 1992). They are principally divided into methods which record and process the speech with one microphone or with two or more microphones.

A series of experiments was carried out using different single microphone processing schemes.

In case of a single microphone, most often the spectral subtraction technique is used, e.g. (Boll, 1979; Vary, 1985). The noise spectrum is estimated during speech pauses. An adaptive filtering is realized with the estimated noise spectrum during speech pauses. The Wiener filter is usually introduced to minimize the average squared error. It turned out that the best results were not obtained using a Wiener filter, i.e. an adaptive filtering based on the short-term power spectra, but using a weighting based on the magnitude of the spectral components (Vary, 1985).

One disadvantage of the spectral subtraction technique is the generation of an artificial noise signal, known as "musical tones". Looking e.g. at a stationary noise segment the short-term spectrum is not a constant in each subband but a random variable distributed around the long-term average. Negative values may occur if a fixed component is subtracted in each subband. These are usually set to zero. The remaining positive values lead to sudden excitations. These can be heard as a splashing noise also called musical tones.

To reduce this artificial noise the magnitude of the negative values is used. A certain amount of the noise which should be reduced remains as a result of this processing method, but the noise level is considerably reduced. The spectral subtraction is applied a second time to this less noisy signal. Further improvements can be achieved by postprocessing methods as proposed in (Boll, 1979) or by elimination of short-term spectral components which are isolated in the time–frequency plane.

This results in a considerable improvement of the signal-to-noise ratio and in a reduction of the musical tones in comparison to a standard spectral subtraction technique.

The short-term energy as well as the zero crossing rate are used for speech pause detection.

The main disadvantages of spectral subtraction are the need for a speech pause detection and the restriction to stationary noise.

These disadvantages can be avoided by an alternative single-channel technique which has shown the ability to considerably improve the recognition rates of an automatic speech recognition system for noisy or reverberant speech (Hirsch, 1992; Hirsch et al., 1991; Hermansky et al., 1991).

Looking at the envelope in a subband reverberation artificially extends the duration of this subband signal with an exponentially decaying characteristic. It turns out that reverberation has a low-pass effect on the envelopes of subband signals (Houtgast et al., 1980). To compensate this effect, an inverse high-pass filter can be applied in each subband to filter the temporal contour of this specific subband. The same constant filter is used in each band. This preprocessing can also be used for noise reduction. Stationary noise results in nearly constant components looking at subband envelopes. Because of this, a significant amount of noise can be removed with this high-pass filtering technique. Also slowly changing noise can be reduced dependent on the frequency characteristic of the high-pass filter. No speech pause detection is required. This is a big advantage in comparison to spectral subtraction.

4. Intelligibility tests

An overview of the test conditions is given in Figure 3. The digitized speech and noise signals can be mixed to realize different signal-to-noise ratios.

The noisy signal is processed with the noise reduction algorithm. The speech is resynthesized

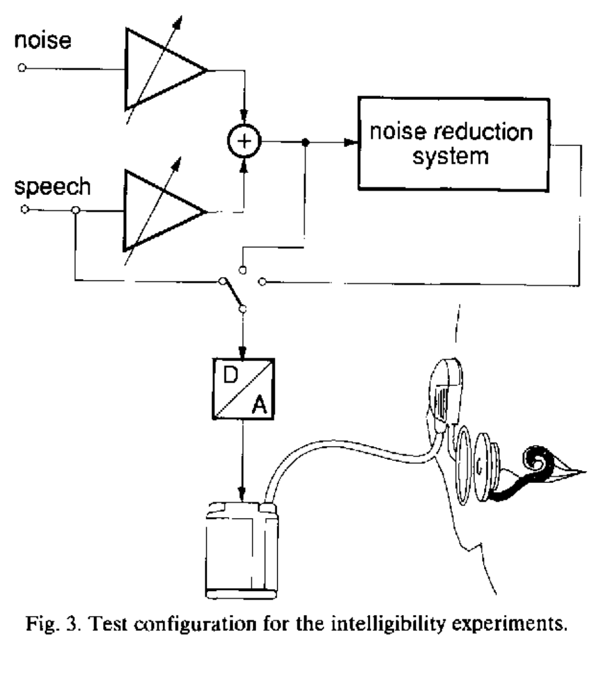

Fig. 3. Test configuration for the intelligibility experiments.

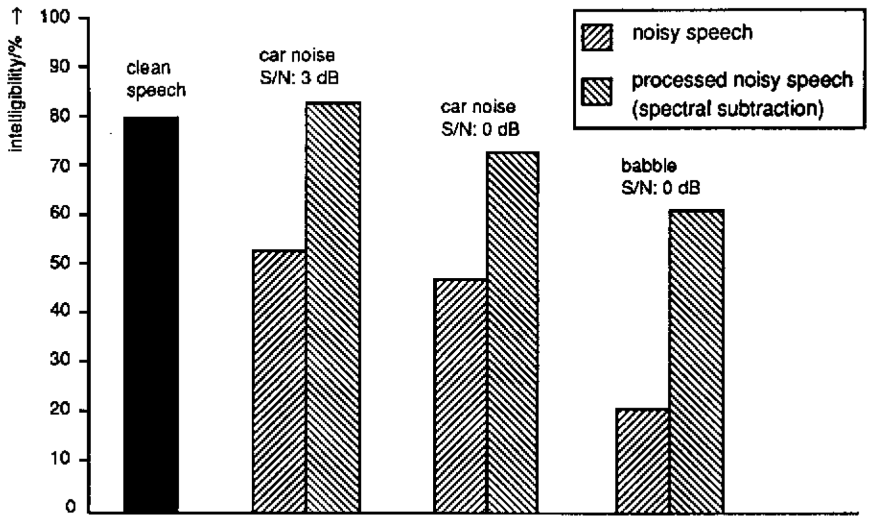

Fig. 4. Speech intelligibility for different noise situations.

with the overlap add method from the processed spectra.

The clean speech, the noisy speech, as well as the processed noisy speech can be fed to the external audio input of a person's cochlear implant system. Twenty-two two-digit numbers were used for the experiments. These numbers are an extract of the so called "Freiburger Sprachtest", a German vocabulary which is often used for intelligibility tests. These speech data are also used for the optimal adjustment of the speech processing unit itself, so that the persons are used to these data. The numbers are randomly presented twice. Some results are given in Figure 4.

Seven persons with an elder version of the speech processor were involved in this experiment who were deaf for a period from 1 year to a couple of years. The intelligibility score of the clean speech was about 80%.

If no noise suppression is applied, the degradation of the intelligibility is significant. Two noise signals were considered, a nearly stationary car noise and a speech babble in a cafeteria, both recorded by ourselves. The car noise was added

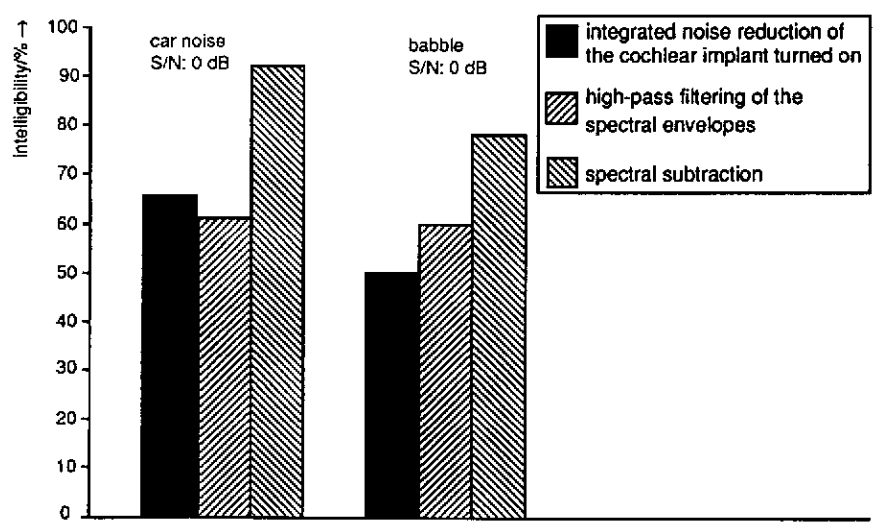

Fig. 5. Speech intelligibility for different noise suppression methods.

at signal-to-noise ratios of 0 and 3 dB, the babble at an SNR of 0 dB.

As expected, the decrease of intelligibility was much higher for the speech like noise. The intelligibility can be improved by processing the noisy speech with the spectral subtraction technique described in the preceding section.

Another small experiment was carried out with two persons with the actual version of the speech processor. This allowed a comparison of the internal noise suppression function of the speech processor with the methods described in the preceding section. Some intelligibility rates are shown in Figure 5.

The high-pass filtering operation of the envelopes in subbands does not significantly improve the intelligibility. A high amount of musical tones is produced by this processing which leads to an unpleasant subjective impression.

This result is in contrast to the automatic speech recognition results. The automatic recognition system does not evaluate the subjective impression, but measures only the spectral distance between clean and processed speech. The spectral distance between the clean and the noisy speech is objectively decreased by the processing.

Best results were again obtained with the spectral subtraction technique.

5. Conclusions and future perspectives

Some experiments showed the possibility of improving the intelligibility of noisy speech for people with cochlear implants. Best results were obtained by using a modified spectral subtraction technique. It turned out that it is better to leave a small amount of background noise rather than to reduce it as much as possible. The latter usually introduces artificial noise like musical tones.

This is one of the rare cases where not only an increasing of the signal-to-noise ratio but also an improvement of intelligibility could be observed for human listeners using a noise suppression technique.

Further investigations will be carried out to improve the intelligibility for these persons in real environments. Possible solutions could be the application of multichannel processing techniques, e.g. (Allen et al., 1977; Martin and Vary, 1992) or a combination of single- and multichannel processing schemes.

A modification of the speech analysis, used in the implant technique up to now, will increase the intelligibility in clean and adverse conditions (Clark et al., 1991; Dillier et al., 1992). The final goal will be an integration of a modified speech analysis and some preprocessing techniques for corrupted speech.

References

J.B. Allen, D.A. Berkley and J. Blauert (1977), "Multimicrophone signal processing technique to remove room reverberation from speech signals", *JASA*, Vol. 62, No. 4, pp. 912–915.

S.F. Boll (1979), "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-27, No. 2, pp. 113–120.

G.M. Clark et al. (1990), *Cochlear Prostheses* (Curchill Livingstone).

G. Clark et al. (1991), "Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant", *J. Octolaryng. Soc. Austral.*, Vol. 5, pp. 354–359.

Cochlear Pty. Ltd. (1989), *Mini System 22 – Audiologist Handbook*.

N. Dillier et al. (1992), "Digital speech processing for cochlear implants", *ORL 54*, pp. 299–307.

ESCA (1992), *Proc. ESCA Workshop on Speech Processing in Adverse Conditions*, Cannes, France.

H. Hermansky, N. Morgan et al. (1991), "Compensation for the effect of the communication channel in auditory-like analysis of speech", *2nd European Conf. Speech Communication and Technology*, Genua, Italia, pp. 1367–1370.

H.G. Hirsch (1992), "Robust Speech Recognition in Noisy and Reverberant Environment", *Nato ASI Series F*, Vol. 75 (Springer, Berlin), pp. 101–106.

H.G. Hirsch, P. Meyer and H.W. Rühl (1991), "Improved speech recognition using high-pass filtering of subband envelopes", *2nd European Conf. Speech Communication and Technology*, Genua, Italia, pp. 413–416.

T. Houtgast, H.J.M. Steeneken and R. Plomp (1980), "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics", *Acustica*, Vol. 46, 1980, pp. 60–72.

J.S. Lim (1983), *Speech Enhancement* (Prentice Hall, Englewood Cliffs, NJ).

R. Martin and P. Vary (1992), "A symmetric two microphone speech enhancement system – theoretical limits and application in a car environment", in *Proc. The 1992 Digital Signal Processing Workshop*, Utica, Illinois, pp. 4.5.1–2.

P. Vary (1985), "Noise suppression by spectral magnitude estimation – Mechanism and theoretical limits", *Signal Processing*, Vol. 8, No. 4, July 1985, pp. 387–400.