Phoneme Compression

Processing of the Speech Signal and Effects on Speech Intelligibility in Hearing-Impaired Listeners

André Goedegebure

Cover: Marieke Goedegebure-Hulshof, André Goedegebure, Infofilm

Printed by Optima Grafische Communicatie, Rotterdam

The research described in this thesis was funded by the European Commission and by the Heinsius Houbolt Fund

Printing of the thesis was supported by Stichting Atze Spoor Fonds and the companies Beltone NV, Oticon Nederland BV, GN Resound BV, Siemens Audiologie Nederland BV, Veenhuis Medical Audio BV

ISBN: 90-8559-052-3

Phoneme Compression: Processing of the Speech Signal and Effects on Speech Intelligibility in Hearing-Impaired Listeners

Foneemcompressie: verandering van het spraaksignaal en effecten op spraakverstaan bij slechthorenden

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 2 juni 2005 om 13.30 uur

door

André Goedegebure
geboren te Duiveland

PROMOTIECOMMISSIE

Promotoren:	Prof.dr. L. Feenstra
	Prof.dr. ir. W.A. Dreschler
Overige leden:	Prof.dr. ir. A.F.W. van der Steen
	Prof.dr. J. Kiessling
	Prof.dr. C.I. de Zeeuw
Copromotor:	Dr. J. Verschuure

“Maar wat hebt u, toen u hier kwam, aangetoond dat u
eerder niet met het gezonde verstand had kunnen aantonen?”

Umberto Eco, uit “Het eiland van de vorige dag”

Voor Marieke

CONTENTS

1	General Introduction	9
2	Compression and its effect on the speech signal	21
3	The effect of compression on speech modulations	43
4	Effects of single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners.	61
5	The effects of phonemic compression and anti-upward-spread-of-masking (anti-USOM) on the perception of articulatory features in hearing-impaired listeners	83
6	Evaluation of phoneme compression schemes designed to compensate for temporal and spectral masking in background noise	107
7	Phoneme compression in experimental hearing aids: effects of everyday-life use on speech intelligibility.	125
8	Final Discussion	145
9	Conclusions	165
	Summary	167
	Samenvatting	171
	List of abbreviations	175
	Dankwoord	177
	Publications	181
	Levensloop	183

1 General Introduction

1.1 Introduction

Hearing aids are commonly used by people suffering from hearing impairment. As indicated by the name, they help the hearing-impaired to achieve a better hearing and speech understanding. However, they cannot fully restore the function of the impaired ear to normal. Hearing-aid users often continue to have problems with poor speech understanding in difficult acoustical conditions. Another generally accounted problem is that certain sounds become too loud whereas other sounds are still not audible. Many hearing-aid users have problems to find the right volume setting in conventional hearing aids. To properly understand a distant speaker they need a high amount of amplification, but this causes environmental sounds and normal speech to become uncomfortably loud. The reason is that the range of levels in between hearing threshold and an uncomfortably loud sensation level has become much smaller due to the hearing impairment. This is often indicated by the reduced dynamic range of the impaired ear.

Dynamic-range compression has been introduced in hearing aids as a possible solution to these problems. After the introduction of digital techniques in hearing aids, it has become a standard processing technique in modern hearing-aid design. Its main function is to provide sufficient amplification at low input levels without overloading the auditory system at high input levels. In this way, listening comfort should improve as the hearing aid compensates for the reduced dynamic range. This effect can already be achieved with a relatively simple slow-acting compression system.

However, the effect of compression on speech and speech intelligibility is still a topic of discussion. It is known that fast-acting compression systems can change the level differences between subsequent speech parts. This means that such a compressor can be considered as a speech processor. Therefore the type of processing is often called syllabic or phoneme compression. From a theoretical point of view, phoneme compression can make weak speech cues audible that otherwise would have been below threshold for the hearing-aid user. However, experimental data do not give a consistent answer to the question whether speech intelligibility improves by using such a device. Many variables play a role, such as the design and the setting of the system, the amount of hearing-loss of the listeners and the acoustical test conditions. With the present thesis I hope to provide a small contribution to the complicated but nevertheless intriguing issue whether signal processing based on phoneme compression can be used to improve speech intelligibility in hearing-impaired listeners.

1.2 Dynamic-range compression

Dynamic range compression is applied in many technical areas, like broadcast and professional audio. The main goal is to reduce the dynamic range of a signal, resulting in a smaller intensity difference between low- and high-level sounds. Although there are various ways to implement such a system, some general principles can be identified. An input-output characteristic defines the relation between the input-level and the corresponding desired output-level. The compression ratio is the slope of this function and determines the theoretical amount of compression applied by the system. Interesting is that both the input and output levels are defined in the logarithmic domain, which is also used by the auditory system. A linear relationship between both logarithmic parameters implies that compression is essentially a *nonlinear* process. A second important relevant parameter is the compression threshold, which determines the active range of the compressor. A first application is the "limiter". This is an example of a compression system with a high compression ratio and a small active range (high compression threshold). It prevents listeners for extreme loud sounds by reducing the amplification of peaks within the signal. Another application is known as Wide-Dynamic-Range Compression (WDRC), using a large active compression range (low compression threshold) and a moderate amount of compression.
The spectral properties of the compressor are determined by the number of independent compression channels. In a multi-channel design the signal is split into different frequency channels and compression is performed within (some of) the separate channels. This results in an effective compression of the signal within the separate frequency bands. The temporal characteristics of the compressor are defined by a set of time constants, consisting of at least an attack time and a release time. The time constants define the time needed by the compressor to realize a change in amplification. The attack time is the time needed to react to a sudden onset and the release time is the time needed to react to a sudden drop in level. When using relatively large time constants, the compressor only reduces differences in overall level. This type of compression is known as Automatic Gain Control (AGC) or Automatic Volume Control (AVC).With short time constants the compressor also reduces the dynamic range of a fast-fluctuating signal like speech. This last type of system is therefore often called a syllabic or (even faster) a phoneme compressor.

The use of digital techniques has considerably increased the possibilities to implement more complex compression systems. This explains a shift from conventional single-channel techniques as compression limiting or AGC towards more sophisticated multi-channel systems as applied in modern hearing aids. Within one hearing aid it is sometimes possible to achieve a very different kind of compression configuration by manipulation of the compression parameters. The choice is up to the clinician. This raises the question what choices should be made for an individual listener. Is there a common rule? What are the key factors that predict the success of some configuration? These are difficult questions that cannot be answered easily. First we have to consider how compression can possibly improve auditory performance in case of a cochlear dysfunction.

1.3 Cochlear compression

When we mention "the ear", most people think about the outer ear (pinna and ear canal) or the middle ear (ear drum and ossicles). Although these parts form an important link in the complete auditory system, the cochlea can be considered as *the* auditory organ. The cochlea transforms the acoustical information into electro-chemical activity of the nerve. It is a complicated organ that is well hidden in the temporal bone and therefore still maintains many secrets about its functioning. The main principle is that a sound wave propagates inside the fluid-filled cochlea and makes the basilar membrane vibrate. Located on the basilar membrane is the organ of corti, which consists of inner and outer hair cells and the tectorial membrane. The hair cells convert the mechanical impulses into electrical energy that the brain can interpret as sound.

One of the most essential cochlear functions in this process is the active outer-haircell system. Due to active mechanical properties of the outer haircells and a complex neural feedback system the cochlea acts as a dynamic-range compressor, also indicated as a compressive nonlinearity. This means that for low- and mid-level sounds the system generates additional electro-mechanical activity that is not present at high-level sounds. As a result the cochlea produces more amplification at low levels than at high levels. This is the main reason why we are able to perceive sounds over a large dynamic range. On the level of the basilar membrane the system acts almost instantaneously. Although some temporal smoothing takes place in the neuronal system, the time constants remain relatively fast (about 10 ms, see Moore and Oxenham, 1998). Furthermore, it is active over a relative broad range (Ruggero and Rich 1991) and can therefore be described as a WDRC system. The compressor is also clearly a multi-frequency system as the cochlea acts as an auditory filter bank with only limited interaction between the different filters.

Unfortunately, the cochlea is a vulnerable system. Although there are many causes of cochlear damage (such as ageing, noise exposure, ototoxic agents, disease, drugs, mechanical disturbance), damage to the outer haircell system is one of the first consequences. Therefore, a cochlear hearing loss implies a poorer functioning of the compressive nonlinearity in most cases. As a result, the quality of the auditory neural information deteriorates. Although the amount of neural activity can be increased by using hearing aids, the loss of quality remains. More stimulation is not per definition better stimulation. Contrarily, the quality may even become poorer at high stimulation levels. It seems therefore logical to compensate for the loss of cochlear compression by using some kind of dynamic-range compression in hearing aids. From a physiological point of view the system should incorporate a wide dynamic range, fast time constants and multiple frequency-channels.

1.4 Psycho-acoustical functions and compression

Auditory dysfunction due to a loss in cochlear nonlinearity can be described in several psycho-acoustical terms, such as a reduced spectral and temporal resolution, a disturbed loudness perception and an increased temporal and spectral masking (Oxenham and Bacon, 2003). They are highly dependent as they all originate from a similar kind of cochlear dysfunction. Signal processing in hearing aids can be used to compensate for these different aspects of reduced auditory functioning.

From the perspective of loudness perception the use of dynamic range compression in hearing aids is almost inevitable. Hearing-impaired listeners with outer-hair cell damage suffer from an increased growth of loudness or loudness recruitment. This means that the loudness increases from soft to loud within a relatively small dynamic range. Dynamic-range compression can be used to map all available input levels within the smaller perceptive range. Loudness discomfort at high levels can be removed by using "simple" compression limiting techniques. More complex systems with multi-channel wide-dynamic range compression are needed to restore loudness perception over a broad range of levels and frequencies (Dillon, 2001). The use of slow time constants is generally sufficient to achieve the desired compensation for loudness deficits, although an additional fast-reacting system may be needed to suppress sudden onsets of high-level sounds (e.g. Moore and Glasberg 1988). In case of fast-fluctuating signals, some evidence has been found that fast-acting compression helps to normalise loudness perception (Wojtczak, 1996).

Another important psychophysical factor that is often related to the use of compression is an increase of temporal masking. Low-level speech parts cannot be perceived due to the presence of preceding high-level sounds. For normal hearing the effect of temporal masking is less pronounced at the mid-levels, which is the active range of the cochlear compression. Therefore, a loss of cochlear nonlinearity results in increased temporal masking effects (Oxenham and Bacon, 2003). In theory dynamic-range compression may help to achieve a release of temporal masking as it emphasises the low-level parts and decreases the influence of the masker. Fast-acting compression is needed as the compression should reduce level differences between the masker and the subsequent masked signal. The results of Moore et al. 2001 suggest that fast-acting compression indeed helps to reduce temporal masking effects on "gap detection" in hearing-impaired listeners.

Next to temporal effects, a reduced spectral processing plays an important role as well. A properly functioning cochlear nonlinearity performs a sharpening of the auditory filters. A loss of nonlinearity results in a broadening of the auditory filters (Patterson and Moore 1988). Therefore, it becomes more difficult for the impaired ear to resolve the spectral information of a broad-band signal and to extract a signal from a background noise. This effect is mostly described as a reduced spectral resolution. Information within a certain frequency band may even be completely masked by high-level signals in adjacent frequency bands. This is known as spectral masking. Because of the asymmetric shape of the auditory filters, there is a higher chance that high-frequency information is masked by low-frequency signals, known as upward-spread-of-masking (USOM). An increased effect of USOM can be shown in hearing impaired listeners by simultaneous stimulation of low- and high-frequency regions (Nelson and Schroder 1996). In general, compression systems are not particularly designed to compensate for a reduced frequency resolution and USOM. Noise suppression and spectral enhancement (Lyzenga et al. 2002) are more appropriate signal-processing techniques for this aim. Nevertheless, fast-acting compression in at least two separate frequency channels may suppress high-level low-frequency sounds and simultaneously increase the level of weak high-frequency cues. This might help to compensate for increased USOM. The use of phoneme compression for this aim will be indicated by “anti-USOM processing”.

1.5 Speech intelligibility

Speech perception in hearing-impaired listeners is hampered by two major factors, deficits in auditory processing as described in the previous section and a loss of audibility. A loss of audibility implies that certain low-level speech sounds cannot be perceived as they are received at sub-threshold levels. A good model has been developed to relate audibility of speech to speech understanding, called the Speech Intelligibility Index (SII, see ANSI S3.79, 1998). This model predicts an almost linear relationship between both factors. A weighing function is used to incorporate the relative contribution of each frequency band to speech perception. Using this model, a loss of audibility of relevant speech signals will generally result in poorer speech intelligibility. Although the SII gives a good prediction on average, individuals may behave differently due to differences in supra-threshold processing capacities. For the more severely impaired ears an extra distortion factor is needed, which is also related to a poor processing quality. Especially for severe high-frequency losses speech intelligibility does not always improve when increasing the audibility at high frequency components (Ching et al. 1998, Turner and Cummings 1999). This implies that next to audibility there is a second important supra-threshold factor that influences speech intelligibility. It is assumed that this factor is related to deficits in auditory processing caused by the cochlear damage.

Dynamic range compression can be used to compensate for each of the two factors. Compression is ideally suited to increase audibility without making the high-level sounds uncomfortably loud. The use of slow time constants is sufficient to restore audibility as long as the dynamic range of the impaired ear is considerably larger than the 30-dB range of speech. Only in case of a smaller residual dynamic range, fast-acting compression may be needed. Next to restoration of audibility, compression can be used to compensate for deficits in auditory speech perception. This is one of the most challenging targets. As mentioned before, such a system is often called a syllabic compressor or a phoneme compressor as it influences the temporal structures within the speech signal. Phoneme compression seems the most appropriate name as the main goal is to affect level differences between subsequent phonemes.

A single-channel phoneme compressor will generally emphasise consonant information and reduce the level of vowels. The reason is that vowels contain on average more energy than consonants, so the compressor tries to reduce these level differences. The level difference between vowel and consonant can be indicated by “the vowel consonant ratio”, which is a more logical expression than the commonly used “consonant vowel ratio”. Decreasing the vowel consonant ratio by phoneme compression of course influences the perception of different speech cues. The aim is to enhance subtle consonant cues, either by achieving a

release of masking from the vowel or by increasing the audibility of spectral cues (Hedrick and Rice 2000). A possible disadvantage is that compression disturbs the perception of the vowel consonant ratio as a speech cue (Hickson and Byrne 1997).
A second approach is the use of multi-channel phoneme compression. This type of compression is better suited to compensate for the reduced auditory functions as the auditory system uses a multi-channel filter-bank as well. Models of loudness (Launer 1995, Moore and Glasberg 1997) or temporal processing (Dau et al 1996) all include independent auditory processing within separate frequency channels. The aim is to improve speech intelligibility by normalising disturbed auditory functions such as loudness recruitment and both increased temporal and spectral masking. A risk of using a high number of independent frequency channels is that spectral contrasts can be considerably reduced. This may have a negative effect on the perception of speech containing spectral cues. Kates (1992) therefore suggests a number of up to 3 channels for optimal compensation of reduced cochlear processing. Another approach is to use some kind of coupling between the various frequency channels.

The most persistent complaints of hearing-aid users are about speech understanding in poor acoustical conditions with reverberation and background noise. Due to the reduced cochlear processing speech cues can no longer be distinguished from the noise signal. For hearing-impaired listeners the “signal-to-noise ratio” (SNR) has to be improved compared to normally-hearing people to achieve the same level of speech intelligibility. Phoneme compression is not the most appropriate technique to improve the SNR. Within each frequency band the noise and the speech are both equally amplified. Between bands the amplification may differ, which can be advantageous if a high-level noise is located in a small frequency area. The compressor will reduce the noise signal within a small band without affecting the speech signal in the other bands. This may improve the overall SNR across channels. Also for noises with a temporally fluctuating character the compressor can selectively reduce the highest noise bursts, which may improve the overall SNR across time. However, for a stationary broadband noise with a speech-like spectrum the effective SNR will remain similar. At high SNR the compressor may even decrease the effective SNR as the low-level noise will be amplified within the speech pauses.
The concept of using compression in such listening conditions is that the listener may still profit from the compensated auditory functions as described in the previous section. The aspect of audibility becomes less important as audibility is mainly determined by the noise level and less by auditory thresholds. Only in a background noise with a fluctuating envelope, phoneme compression can improve audibility by increasing the speech levels within the temporal gaps of the noise. Also, a release of temporal masking may be expected when using phoneme compression in this type of background noise.

There is also a risk of obtaining adverse effects from fast-acting compression in background noise. The Speech Transmission Index (STI) is a good model to predict speech intelligibility in background noise. It is based on the amount of modulations that is left within the various frequency channels. The higher the amount of modulations, the more speech information is available. Multi-channel phoneme compression will always reduce the amount of modulations and therefore apparently reduce the available speech information as well. This classical argument against the use of phoneme compression is described by Plomp (1988). The counter-argument of Vilchur (1989) was that the STI model modulations are reduced by adding stationary noise to the speech signal, thus the low-level speech parts are not available anymore. With compression all speech parts are still available, only the level differences are changed. This means that the STI cannot be applied to compressed speech, which is experimentally confirmed by (Kollmeier and Hohmann, 1995). However, there is some consensus that a high amount of modulation reduction may decrease speech intelligibility because the modulations cannot be detected anymore. The highest chance for such adverse effects on intelligibility can be expected when using high amounts of compression in stationary background noise conditions. Due to the background noise, the amount of modulations is limited. After a substantial *additional* reduction of modulations by the compressor, some temporal information may not be available anymore to the listener.

1.6 Performance with compression

Many studies have evaluated the effect of phoneme compression in hearing-impaired listeners. Unfortunately, the number of studies that show no or negative effect of phoneme compression exceeds the number of studies that report positive effects. In quiet, mostly no effect has been found on speech intelligibility compared to a well-defined linear reference at comfortable presentation levels (Dillon , 2001). Verschuure et al. (1993) found small positive effects on consonant perceptions in listeners with moderate-to-severe high-frequency losses, using phoneme compression in one high-frequency channel. Other studies found positive effects at input levels below comfortable levels (Bustamente and Braida 1987, Vilchur 1987). In these studies similar positive effects might be expected for slow-acting compression systems as restoring audibility plays a major role at low input levels. The relation between audibility and performance with fast-acting compression is nicely illustrated by Souza and Bishop (2001). They found that the use of fast-acting compression may have an additional value at suboptimal input levels for hearing-impaired listeners with severe high-frequency losses. At levels of maximum audibility a similar performance was found with both linear amplification and fast-acting compression. As long as audibility is optimised with linear processing it seems difficult to obtain additional benefit from phoneme compression.
In background-noise conditions there are a few studies that show a positive effect of phoneme compression. Yund and Buckles (1995a) found improvements in a stationary speech-shaped background noise with a multi-channel compression system. Moore and Glasberg (1998) also found a positive effect in stationary background noise, using a compressor with only two channels. However, many other studies did not found improvements in stationary noise (e.g. Bentler and Nelson 1997, Franck et al. 1999, de Gennaro et al. 1986, van Harten-de Bruijn et al. 1997, Kollmeier et al. 1993, Marzinzik et al. 1997, Moore et al. 2004, Olsen (2004), Vilchur 1987, Walker et al. 1984) or even found considerable negative effects (Drullman and Smoorenburg 1997, van Buuren et al. 1999). In background noises with a fluctuating

envelope positive effects were found in some studies (Moore et al. 1999, Verschuure et al. 1998), whereas others did not find any improvement (van Buuren et al. 1999, Franck et al. 1999, van Harten-de Bruijn et al 1997, Olsen 2004).

In view of these results, no consistent preference is shown for a specific design of phoneme compression. Studies using compression systems with many frequency channels do not reveal substantially better results than the studies using systems with only a limited number of channels. Some results indicate that the performance in background noise may slightly improve using a larger number channels (Moore et al. 1999 and Yund and Buckles 1995b). Furthermore, fast time constants are needed to achieve any effects on speech intelligibility. Studies that report significant positive or negative effects typically use time constants of 1-10 ms for the attack time and 10 -50 ms for the release time. The compression ratios used in the various studies typically varied between 2 and 8. A negative effect of increasing the compression ratio in background noise was reported by numerous studies (van Buuren et al. 1999, Moore et al. 1992, Olsen 2004). This effect has been related to the loss of temporal information due to an extreme reduction of the available modulations (Plomp 1988, see also previous section). The use of moderate values of the compression ratio (2-3) seems therefore to be preferable over the use of high values (>4), especially in conditions with stationary background noise.

1.7 Aims of the study

The research presented here was mainly performed within the framework of the European projects HEARDIP and SPACE. The aim of these projects was to develop hearing-aid strategies to improve speech intelligibility in hearing-impaired listeners. We focussed on the effects of phoneme compression, using a two-channel approach that showed to be promising in earlier studies (Verschuure et al. 1994, 1998). The concept is based on enhancing high-frequency speech cues without introducing undesirable side effects on the spectral fine-structure of the signal. The balance between low- and high-frequency gains is continuously changed dependent on the input level of each phoneme. It is mainly designed for hearing-impaired listeners with moderate-to-severe high-frequency hearing losses. Therefore we have evaluated the effects of compression only for this subgroup of hearing-impaired listeners. The main research questions are:

- What effect have different types of phoneme compression on speech and speech modulations (chapters 2 and 3)
- What is the effect of our type of processing on speech intelligibility in hearing-impaired listeners (chapters 4 and 6)
- What are the effects of small but fundamental changes in our configuration on speech intelligibility (chapters 4 and 6)
- Can changes in perceptual strategies be identified that explain the effects of speech processing? (chapter 5)
- Is the system appropriate for use in everyday environments? (chapter 7)

- Does regular use of the speech processing have an effect on performance? (chapter 7)

1.8 References

ANSI S3.79 (1998) American National Standard Methods for the calculation of the speech intelligibility index. American National Standards Institute, New York.

Bentler RA, Nelson JA (1997) Assessing release-time options in a two-channel AGC hearing aid. J Am Acad Audiol 6:43-51.

van Buuren RA, Festen JM, Houtgast T. (1999) Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. J Acoust Soc Am.105(5):2903-13.

Bustamente DK, Braida LD (1987) Prinicipal-component amplitude compression for the hearing impaired. J. Acoust Soc Am 82(4): 1227-1242.

Ching TY, Dillon H, Byrne D (1998) Speech recognition of hearing-impaired listeners: predictions from audibility and the limited role of high-frequency amplification. J Acoust Soc Am. 103(2):1128-40.

Dau T, Puschel D, Kohlrausch A (1996). A quantitative model of the effective signal processing in the auditory system. I. Model structure. J. Acoust Soc Am 99(6), 3615-3622.

Dillon H. (2001) Hearing aids. Sydney: Boomerang Press.

Drullman R, Smoorenburg GF. (1997) Audio-visual perception of compressed speech by profoundly hearing-impaired subjects. Audiology. 36(3):165-77.

Franck BAM, Kreveld-Bos CSGM, Dreschler WA, Verschuure J (1999) Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. J. Acoust Soc Am 106 (3):1452-1464

van Harten-de Bruijn HE, van Kreveld-Bos CSGM, Dreschler WA, Verschuure J. Design of two syllabic non-linear multi-channel signal processors and the results of speech tests in noise. Ear Hear 1997; 18:26-33.

Kates JM (1993) Toward a theory of optimal hearing aid processing. J Rehab Res Dev 30(1): 39-48.

Kollmeier B, Peissig J, Hohmann V. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. J Rehab Res Dev 1993; 30:82-94.

Launer S (1995) Loudness perception in listeners with sensorineural hearing impairment. PhD thesis, Universität Oldenburg.

Lyzenga J, Festen JM, Houtgast T (2002) Speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity. J Acoust Soc Am 112(3): 1145-1157.

Marzinzik M, Hohmann V, Appel JE, Kollmeier B. (1997) Evaluation of different multi-channel dynamic compression algorithms with regard to recruitment compensation, quality and speech intelligibility. in Seventh Oldenburg symposium on psychological acoustics. Oldenburg.

Moore BCJ, Glasberg BR. (1988) A comparison of four methods of implementing automatic gain control (ACG) in hearing aids. Brit J Audiol 22:93-104.

Moore BCJ, Peters RW, Stone MA. (1999) Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J Acoust Soc Am 105 (1): 400-411.

Moore BCJ, Glasberg BR, Alcantara JI, Launer S, Kuehnel V. (2001) Effects of slow- and fast-acting compression on the detection of gaps in narrow band noise. Br J Audiol 35:365-374.

Nelson DA, Schroder AC. (1996) Release from upward spread of masking in regions of high-frequency hearing loss. J Acoust Soc Am. 100(4):2266-77.

Olsen HL (2004) Supra-threshold hearing loss and wide dynamic range compression. Thesis at the Karolinska Institutet Stocholm, published by Elanders Gotab, ISBN 91-7349-921-8.

Oxenham AJ, Bacon SP (2003) Cochlear compression: perceptual measures and implications for normal and impaired hearing. Ear Hear. 24(5):352-66.

Plomp R (1989) The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function. J Acoust Soc Am 83: (2322-2327).

Patterson RD, Moore BCJ (1986). Auditory filters and excitation patterns as representations of frequency resolution. In Moore BCJ(Ed.) Frequency selectivity in hearing (123-177) London: Academic Press.

Ruggero MA, Rich NC (1991) Furosemide alters organ of corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. J Neurosci 11(4): 1057-1067.

Souza PE, Bishop R (2000) Improving audibility with nonlinear amplification for listeners with high-frequency loss. J Am Acad Audiol 11:214-223.

Turner CW, Cummings KJ. (1999) Speech audibility for listeners with high-frequency hearing loss. Am J Audiol. 8(1):47-56.

Villchur E. (1987) Multiband compression processing for profound deafness. J Rehabil Res Dev 24:135-148.

Villchur (1989) Comments on " The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function". J Acoust Soc Am 86 (1): 425-427.

Verschuure J, Dreschler WA, de Haan EH, van Cappellen M, Hammerschlag R, Mare MJ, Maas AJ, Hijmans AC. (1993) Syllabic compression and speech intelligibility in hearing impaired listeners. Scand Audiol Suppl.38:92-100.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP. (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Walker G, Byrne D, Dillon H. (1984) The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. J Acoust Soc Am 76(3):746-757.

Wojtczak M (1996). Perception of intensity and frequency modulation in people with normal and impaired hearing. In B. Kollmeier (Ed.), Psychoacoustics, speech and hearing aids (35-38). Singapore: World Scientific Publishing Co.

Yund EW, Buckles KM. (1995a) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239.

Yund EW, Buckles KM. (1995b) Multichannel compression hearing aids: effect of number of channels on speech discrimination in noise. J Acoust Soc Am. 97(2):1206-23.

2 Compression and its effect on the speech signal

Based on Ear&Hear (1996) 17(2):162-175.
Verschuure J, Maas AJJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA

Abstract

Compression systems are often used in hearing aids to increase the wearing comfort. A great deal of attention has been given to the static parameters but very little to the dynamic parameters. We present a general method to describe the dynamic behaviour of a compression system by comparing modulations at the output with modulations at the input. The use of this method is described for an experimental digital compressor developed by the authors, and the effects of some temporal parameters such as attack and release time are studied. This method shows the rather large effects of some of the parameters on the effectiveness of a compressor on speech. The method is also used to analyze two generally accepted compression systems in hearing aids. The theoretical method is next compared to the effects of compression on the distribution of the amplitude envelope of running speech, and it could be shown that single-channel compression systems do not reduce the distribution width of speech filtered in frequency bands. This finding questions the use of such compression systems for fitting the speech banana in the dynamic hearing range of impaired listeners.

2.1 Introduction

Many patients complain of problems with hearing resulting from reduced dynamic range. Listening to a discussion involving multiple speakers, each talking at a different level, or listening under different acoustical conditions requires patients frequently to readjust the volume control on their hearing aids. In some patients with large losses, and thus small dynamic hearing range, even the dynamics of the speech signal itself cause problems; amplifying the weak parts of the speech to audible levels causes the strong parts to be uncomfortably loud.

Compression systems have been used to help patients in this respect for many years. Recently, the number of available automatic hearing aids has surged with a number of different algorithms of automatic gain control (AGC). The algorithms are defined by many parameters of which the resulting effect on the speech signal is not always clear. As speech is a fluctuating speech signal, the choice of the time constants determines whether the dynamics of the speech signal are reduced by the compressor. With relatively slow time constants the compressor only corrects for the overall level, while with fast time constants the level differences between consequent speech parts are affected. There is however no clear definition of how fast or slow time constants should be to achieve a desired effect on the speech signal. Furthermore, the choice of frequency channels and the static compression ratio (CR) will also influence the resulting effect of compression on speech.

It is the purpose of this chapter to analyze what a compression system does to a speech signal, and in particular to the speech parts relevant for speech recognition, and to specify the parameters of a compression system that are relevant for different compression goals. Our approach involves a theoretical method to measure the effectiveness of compression systems and the comparison of the outcome measure of this measurement with that of the level distribution of a "normal" speech signal.

2.2 Description of speech and compression systems

Speech

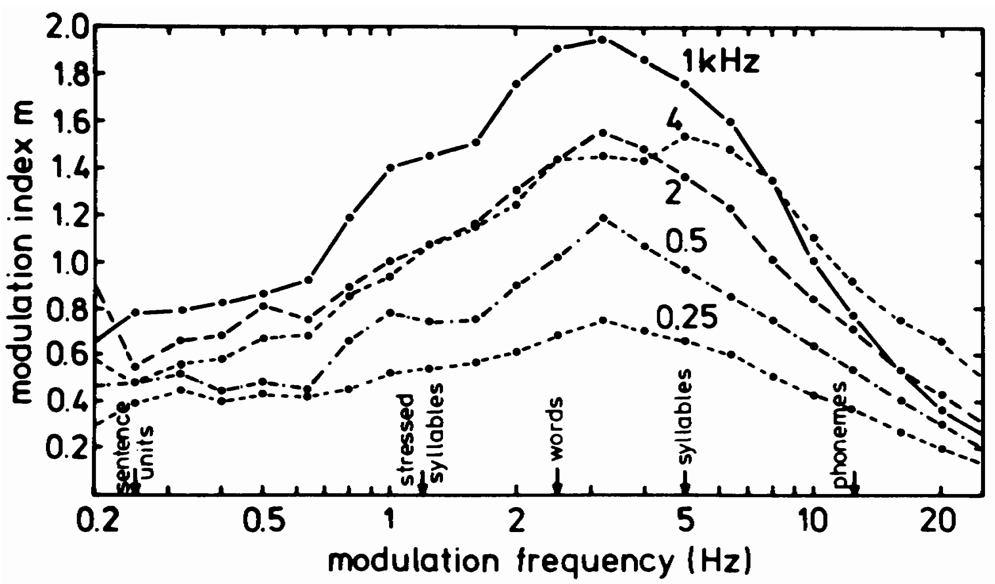


Figure 2-1 *Modulation spectrum of running speech per octave frequency band (3). From Plomp (1983) with permission.*

A speech signal can be described in physical terms as a modulated spectrum. Both aspects, spectral information and temporal information, are relevant. The two aspects are not equally important for all parts of speech (Verschuure et al., 1993). Vowels, semivowels, and nasals require good spectral resolution (separate detection of *F1* and *F2)* while there is little information in the (almost absent) modulations. Fricatives and plosives, on the other hand, are strongly modulated signals differing mainly in time structure (e.g., the gap before the plosive) while only crude spectral analysis is required. In hearing-impaired individuals, frequency resolution is reduced through broadening of the auditive filters and excessive upward spread of masking. The temporal resolution is reduced because speech is presented closer to the threshold where the effective temporal resolution is poorer than at higher levels. These facts and the possible consequences for speech recognition have been discussed by Verschuure et al. (1993).

The average spectrum of speech seems well established, even as a function of level (Pavlovic, 1992). It can be described roughly as having a peak around 400 Hz and falling off above 500 Hz at a rate of about 10 dB/oct. The modulations can also be represented by a spectrum, the amplitude-modulation spectrum (Plomp, 1983). The relevant frequencies of this spectrum range roughly from 0. 1 to 40 Hz. The modulation spectrum can be determined per octave band of the speech spectrum. Such an analysis shows that the modulation spectrum hardly depends on the octave band of the speech spectrum from which it is taken. The frequency of maximum modulation is around 3 Hz, and the maximum amount of modulation is found in the frequency band around 1 kHz (figure 2-1). For the high-frequency band (4 kHz), the maximum shifts somewhat toward a higher modulation frequency (5 Hz). The figure also shows the modulation frequency that can be associated with phonetic entities. The stress pattern has a modulation frequency of up to about 1 Hz, words cause a modulation of around 2.5 Hz, syllables of around 5 Hz, and phonemes of around 12 Hz. The detection of the silent gap in a plosive (duration 30 to 60 ms) requires an even higher modulation frequency (up to about 30 Hz).

Compression System

Compression systems are mainly developed to compensate for the reduced dynamic range of the hearing impaired listener. Both spectral and temporal patterns of speech are adjusted by using compression. The combination of compression parameters defines to what extent the speech signal is altered. Compression systems in hearing aids can be characterized in terms of a number of parameters (ANSI S3.22.1987):

1. Compression ratio. This measure represents the effectiveness of the compressor: in hearing aids it is usually given as the ratio between a rise in input and the resulting rise in output level (both in dB). A compression ratio of 4 means that the output is raised by 1 dB for each rise in input level of 4 dB.
2. Bandwidth. This measure specifies the frequency range over which the compressor is active.
3. Number of bands. This measure specifies the number of frequency bands into which the signal is split. Each frequency band may have different compression parameters.
4. Compression threshold. This is the knee-point of compression, giving the input or output level below or above which a particular compression circuit is activated.
5. Control signal. This measure is derived from either the input or the output signal, and it describes the feedback signal that controls the compression.
6. Attack time. The attack time is a measure of the speed with which the amplification is adjusted after a rise in input signal level. It may be specified as a 1/e, 10 to 90% decay time, or any other time constant. For hearing aids, attack time is defined as the time required for the output of a hearing aid to reach 2 dB of the steady-state level after a sudden increase in level from 55 to 80 dB SPL when stimulated by a sine wave of 2 kHz.

7. Release time. Release time is a measure of the speed with which the amplification is adjusted after a drop in input signal level. It can also be specified as a 1/e, 10 to 90% decay time, or as the time required for a 20 dB drop from a certain level. It is defined for hearing aids as the time required for the output of a hearing aid to reach 2 dB of the steady-state level after a 2 kHz sine wave is decreased in level from 80 to 55 dB SPL.
8. Delay time. In the compression system with suppressed overshoots that was designed by Verschuure et al (1993) and Verschuure et al (1994) this is one additional parameter to be dealt with. This parameter specifies the delay in milliseconds between the speech input signal and the compression control signal. Such a delay may be used to anticipate events in the signal such as jumps in level.

It is clear that a great number of parameters are required to characterize the function of a compression system. Knowledge of these parameters is insufficient to predict how the system will affect the speech signal. For steady-state signals, only spectral effects are important, and the spectral effects can be computed from the input spectrum, the compression threshold, and the compression ratio in each frequency band. However, speech is a modulated signal, and temporal effects should also be taken into account.

2.3 Method 1: effective compression of a modulated signal

Theoretical base line

A compression system is designed to reduce modulations to the extent given by the compression ratio. The effectiveness in modulation terms can be given by modulation frequencies for which the modulations are effectively reduced and by modulation frequencies for which the reduction fails. The effectiveness can be measured by determining the modulation of a signal at the output given a certain modulation at the input (Verschuure et al 1992). The method is similar to the one independently developed by Stone and Moore (1992). The approach has some relation to the approach used by Steeneken and Houtgast (1980) for measuring speech intelligibility in transmission lines, except that it now is used to measure the desired modulation reductions produced by hearing aids.

The method uses a sine wave with carrier frequency (ω_c) and a modulating frequency (ω_m). The amplitude of the carrier is a_0 and the amount of modulation is given by the modulation index m. Comparison of the modulation index at input and output of the compressor gives a measure of the effectiveness of the compressor. The amount of modulation can be determined by accurately analyzing the spectrum of the signal. The theoretical issues are presented in the appendix (section 2.10). Here only the main points necessary to understand the method are presented.

An amplitude modulated sine wave can be described by:

$$x(t) = a_0(1 + m\cos(\omega_m t))\cos(\omega_c t) \qquad (2.1)$$

The spectrum of the signal consists of a carrier and two side bands. The difference in level between the two side bands and the carrier is in dB:

$$\Delta S = 20 \ \log m - 6 \tag{2.2}$$

Equation 2.2 shows that the comparison of the spectral level of the central component with that of the side bands provides a direct measure of the modulation depth. We now can determine the spectral level difference of a modulated sine wave at the input (ΔS_i) and at the output (ΔS_o). From these values we can compute the effective compression ratio CR_{eff} as:

$$CR_{eff} = 10^{\frac{\Delta S_o - \Delta S_i}{20}} \tag{2.3}$$

The derivation of this equation is shown in the appendix, as well as a limitation to the method. The compressor works in the logarithmic (dB) domain, whereas spectra are determined as linear measures. A sinusoidal modulation is distorted by a logarithmic compressor. The distortion shows up in the spectrum as higher-order side bands and extra contributions to the carrier and first-order side bands. The distortions are small and negligible for small modulation depths, and for m < 0.45 the errors are smaller than 5%. For this value the harmonics of the modulation side bands are more than 40 dB down for the second harmonic and more than 50 dB down for the third harmonic. On the other hand, for these modulation depths, the differences in level can be determined with a high enough accuracy to warrant the use of the method.

Experimental compressor

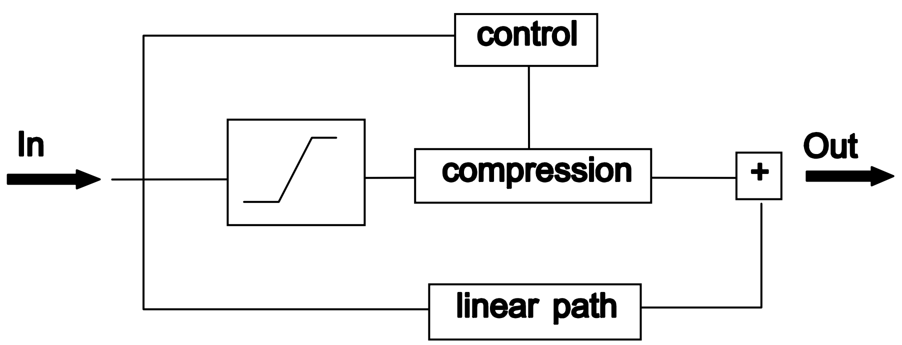


Figure 2-2. *Block diagram of the smoothed compressor implemented on a DSP 56001.*

We used a compressor implemented on a digital signal processor DSP 56001 for testing the method. The design has been described in detail by Verschuure and Dreschler (1993), Verschuure et al (1993, 1994) and Goedegebure et al 2000. The block diagram is shown in figure 2-2. The principle of the design is a two-channel processor. The compressor works only on the signal in the second channel, which is a filtered part of the speech signal (FIR-filter), usually the high-frequency part of the signal. The filter shape is computed as the inverse of the hearing loss. The signal passing through the filter is then delayed. The compressor control signal is taken directly from the input signal (it may be additionally filtered) and rectified. The rectified amplitude signal is compared with a table to determine the required amplification of the signal passing through the FIR-filter and delay. The signal passing through the unfiltered linear path is added to the compressed and filtered signal after a proper delay and in such a way that the frequency response of the total system for a signal presented near the level of maximum intelligibility is a half-gain response. The design is made in such a way that this is also the response for linear amplification. A stylized frequency response is shown in figure 2-3.

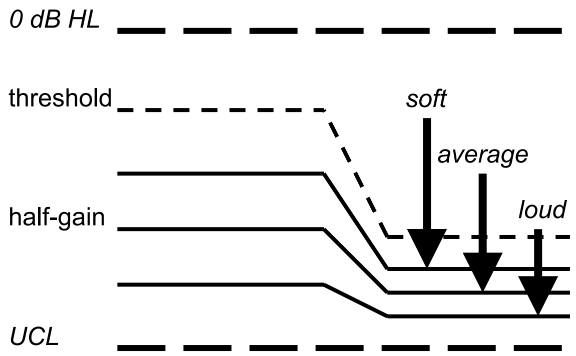


Figure 2-3 *Stylised frequency responses with smoothed compression as a function of level. The half-gain response coincides for all compression conditions including linear amplification.*

A delay is used to suppress overshoots at the onset of a louder part of the signal and a peak-hold circuitry is used to suppress overshoots at the offset of the signal. The action of these features results in a system in which the higher-level parts show no overshoot (temporal distortion). All temporal distortion is transferred to the low-level parts of the signal where it most probably will not be detectable for a hearing impaired person because of poorer temporal resolution and threshold effects (Verschuure et al, 1993).
The compression ratios of the implemented system can be set at 1 (linear), 2, 4, and 8. If not stated otherwise, the attack time was set at 5 ms, the release time at 15 ms, the delay time at 3 ms, and the compression threshold at -70 dB versus maximum input level. The static input-output characteristics showed the compression ratio to be correct over a range of at least 65 dB. Speech was presented at a root-mean-square level of -22 dB versus maximum input level, leaving enough head room to avoid high distortion. The effects of the compressor and the choice of parameters on speech intelligibility have been described elsewhere (Verschuure and Dreschler 1993, Verschuure et al 1993, Verschuure et al 1994) and will not be repeated here.

Measurement Procedure

We used a Hewlett Packard Dynamic signal analyzer type 35665A for the analysis. The span was adjusted to 50 Hz or 100 Hz with a resolution of 400 lines. The dynamic range in fast Fourier transform mode was 72 dB. We analyzed the effectiveness of the compressor for modulation frequencies of 2, 4, 8, 16, 24, 32, 40, and 50 Hz.

2.4 Results Method 1

Modulation depth of test signal

This method of spectral analysis can be used effectively only if the test signal has a small modulation depth (see appendix, section 2.10). We wanted to check the theory by determining the effective compression ratio for various modulation depths. The carrier frequency was set at 2 kHz, a frequency in the middle of the band where the compressor was active, and the compression ratio was set at 4. In figure 2-4 we show an example of a set of curves determined for modulation depths of 20 (squares), 40 (triangles down), 60 (dots), and 90% (triangles up). The theory predicts that only the smallest two modulation depths give reliable results within a 5% error (absolute error of 0.2). The point at 0 Hz was taken from the static compression curve.

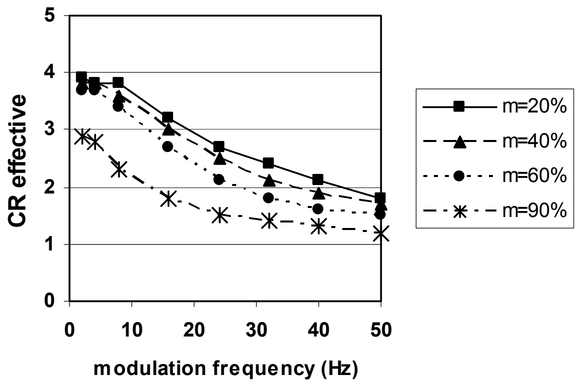


Figure 2-4 *Effect of modulation depth on measurement of effective compression. The modulation depths are 20% (squares), 40% (triangles), 60% (dots) and 90% (stars).*

We saw that the difference between the 20% and the 40% curve was within the margin of error and that the other curves deviated more, as was expected. We concluded that the method works well within the theoretical limitations of using modulation depths smaller than 45%. In all further experiments, we used a modulation depth of around 40%. Note that speech involves much higher modulation depths than 45%. The consequences of this finding will be discussed later on.

For the modulation depths below 45% the compressor is effective for a broad range of modulation frequencies. As we find a kind of low-pass filter response, we define a cut-off modulation frequency as the half-value (-3 dB) point. Figure 2-4 shows that the effective modulation bandwidth of this compressor is about 40 Hz. The cut-off modulation frequency may be interpreted in terms of its effect on speech features. Comparison with figure 2-1 suggests that the whole speech signal is effectively compressed with this compressor. In terms of time constants of signals, this value can be interpreted as reducing changes between signal parts occurring within 25 ms. Such a speed is enough to make the compressor adjust its amplification within the silent gap before a plosive. We shall use the term phonemic compressor for such a fast compressing system.

Release Time of Compressor

The attack and the release times of the compressor have a pronounced effect on the effectiveness of the compressor. In hearing aids, the attack time is usually short, but the release time in commercial hearing aids shows large differences from values of about 70 ms up to 150 ms, typically around 125 ms.

We measured the effective compression ratio as a function of the attack and the release time for a 2-kHz carrier frequency in the middle of the frequency band in which the compressor is effective. The level was set to the root-mean-square level of speech (-22 dB versus maximum level). The modulation depth was 40%. The effect of varying the release time is shown in Figure 2-5 for an attack time of 5 ms. A similar result was obtained from varying the attack time.

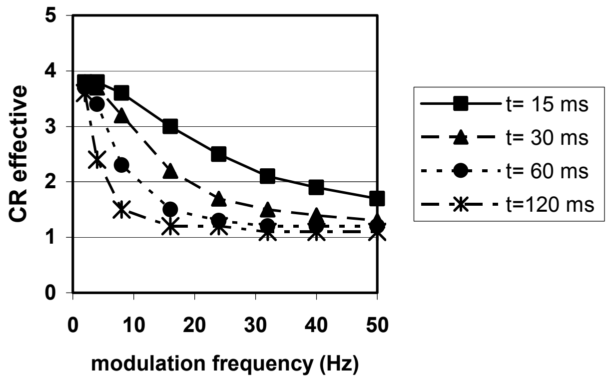


Figure 2-5 *Effect of release time on effective compression ratio. The release times are 15 ms (squares), 30 ms (triangles), 60 ms (dots) and 120 ms (stars).*

We see the expected reduction of effectiveness of the compressing system. The cut-off modulation frequency changes from about 40 Hz to 20, 10, and 5 Hz. In terms of speech, the reduction of the cut-off modulation frequency means that for a release time of about 100 ms, as is often used in hearing aids, only differences between words, not between syllables, are compressed. It can be interpreted that only the level variations present in the intonation pattern are effectively filtered out, but that the amplitude distribution would hardly be affected.

Effect of smoothing the amplitude envelope by delay

Verschuure et al. (1993) and Verschuure and Dreschler (1993) argue that overshoots can be expected to reduce the effectiveness of the compressing system quite drastically. If a change in level occurs (e.g., at vowel onset or offset, at onset of plosive, etc.), the system will transmit the jump in level without any reduction, even if the system is fast. To make the system more effective, we introduced smoothing of the amplitude envelope by time delay and peak hold. The effect of delay on the effective compression ratio was tested next. The conditions were a 2 kHz carrier modulated at 45% and set at a level 22 dB below the maximum level (the RMS level of the speech signal) with an attack time of 5 ms and a release time of 15 ms. Figure 2-6 shows the determined effective compression ratio for various delays. It is striking to see that a delay of just 3 ms increases the effectiveness of the compressor by a factor of about 2. The cut-off frequency moves up from about 18 Hz to 35

Hz. A delay that is too long is counterproductive and even leads to an expansion of faster modulations (above 30 Hz).

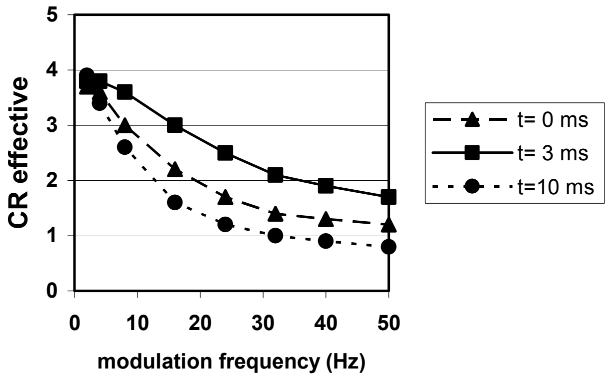


Figure 2-6 *Effect of delay time on the effectiveness of the compressor. The delay times are 0 ms (triangles), 3 ms (squares) and 10 ms (dots).*

Commercially available hearing aids

We tested a Philips S45-1 hearing aid to represent a conventional straightforward AGC hearing aid. The parameters of the Philips hearing aid were a compression ratio of 5 for input levels above 65 dB SPL, an attack time of 5 ms, and a release time of 110 ms. The static responses were checked and found to be correct. The effective compression ratio of the S45-I is shown in figure 2-7. The figure shows that the compression system has a cut-off modulation frequency of less than 4 Hz. This indicates that the hearing aid only suppresses the level information on the intonation pattern and slower changes in level. The system is not effective at evening out differences in level between words.

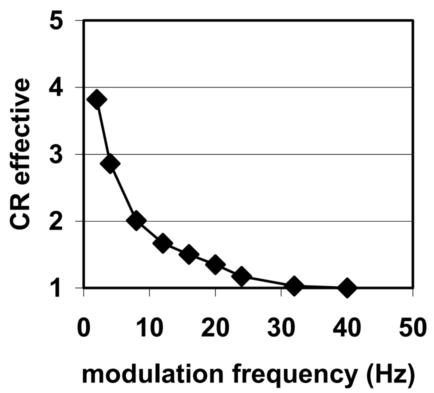


Figure 2-7 *Effective compression ratio of typical commercial hearing aid (Philips S45-I).*

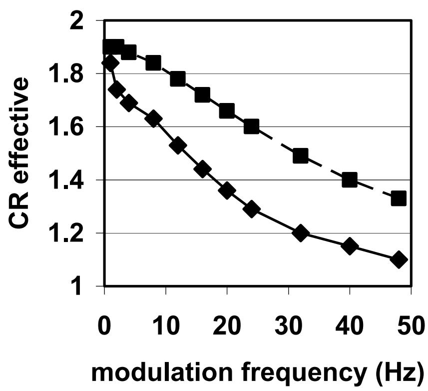


Figure 2-8 *Effective compression ratio of K-amp (squares) as compared to that of the experimental compressor (diamonds).*

Next we tested a K-amp amplifier. The K-amp has a level-dependent frequency response that is similar to that of our experimental compression system. The compression system is an adaptive compression circuit. The compression ratio is just above 2. The attack time is 5 ms, the release time depends on the amount of time a signal has been on. Killion (1993) stated that the K-amp's release time for short signals such as slamming doors is around 20 ms; the release time rises sharply for signals longer in duration than about 100 ms, such as speech, to a release time of about 600 ms for signals longer than about 1 sec. Without the modulation analysis approach used here, it is difficult to determine how highly modulated signals such as speech are affected by such complex processing systems.

The effective compression ratio of the K-amp is shown in figure 2-8 (as diamonds) together with the effectiveness of our experimental compressor (squares) for compression ratio 2. Figure 2-8 shows that the cut-off modulation frequency of the K-amp is about 12 Hz, which means that the compressor is fast enough to influence relevant modulation frequencies of speech. On the other hand, given the similarity in time constants of the K-amp and our smoothed compressor, the K-amp is less effective. If we take the effect of smoothing into account (figure 2-6), the difference can be understood to be the result of a small difference in release time and of the overshoots. Interpreting the data in terms of speech modulations, the K-amp seemed to be effective in evening out differences between words and syllables but did not seem to be fast enough to enhance consonant recognition such as the detection of a gap before a plosive. The short time constant may further cause overshoots, leading to false articulation cues.

Conclusion

We conclude that the method distinguishes between different types of compressors and different settings of the time constants. It provides us relevant information about the temporal behaviour of the compressor at a certain frequency. The translation to speech signals is however not straightforward as speech is a broad-band signal with modulations in different frequency bands. Furthermore, the modulation depth of speech exceeds a value of 45% that is recommended as a maximum in the present method. Therefore a second method will be introduced using running speech as test signal.

2.5 Method 2: compressing amplitude distributions of speech signal

All of our remarks about the effectiveness of compressors were based on the comparison of the effective amplitude cut-off frequency for compression and the prevalent modulations in speech (figure 2-l). It is necessary to measure the effect of compressors on real speech.

Uncompressed Speech

For this analysis, we used a 32 sec sample of recorded continuous speech without pauses or hesitations spoken by one Dutch actor. The recording was made on DAT-tape by a radio engineer of the Dutch Broadcasting Enterprises without using any form of compression. The actor has a general accent and good articulation and intonation.

The RMS level was adjusted to the -22 dB versus maximum level of our compressor. To diminish the background noise, the compression threshold was set at -45 dB. The speech signal was rectified, and the amplitude was sampled with a cut-off frequency of 50 Hz at a rate of 1024 samples per second by the Hewlett Packard Dynamic signal analyzer type 35665A. Linear amplitudes were transformed in dB and counted in bins using a spread-sheet program. An effective compression ratio was calculated from the level distributions by dividing the estimated width with and without compression. The width was estimated at a level of 50% from the peak value of the level distribution.

Wide-Band Control

Figure 2-9 shows the results of analyzing the output from the smoothed phonemic compressor as shown in figure 2-2 using the flattest and widest filter band possible. The dB-scale is an arbitrary scale of which 68 dB coincides with the RMS level of unprocessed speech and 45 dB with the compression threshold of the compressor. The four line types are for compression ratio 1 (linear; solid line), 2 (dashed), 4 (dash-dotted), and 8 (dotted). To compare the widths of the patterns, the number of counts was scaled for the peak values to coincide; no shifts in the level direction were performed.

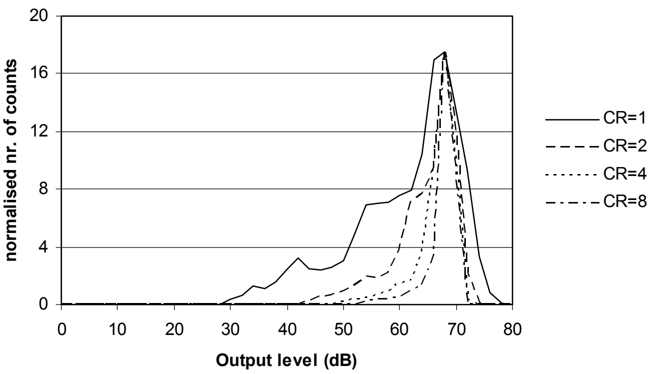


Figure 2-9 *Level distribution of continuous speech for linear amplification (drawn line) and compression ratio 2 (dash), 4 (dot) and 8 (dash-dot).*

The width of the uncompressed speech was about 45 dB, not uncommon for normally intonated speech. The compressor reduced this width substantially. If we analyze the width of the distribution pattern at a count number 20 dB below the peak value (0. 1 x maximum number), we get for the compression ratios 2, 4, and 8 -a reduction of the width by a factor of 1.9, 3.8, and 5.2. The compressor is indeed reducing the distribution of speech amplitudes.

We also analyzed the level distribution of the same segment of speech in octave bands. The compressor control was the same as wide-band control. The only difference compared with figure 2-9 is that now only the distribution of signal levels in the 2 kHz octave band are analyzed. The results are given in figure 2-10. We see that compression results in a little more gain for the high frequencies for higher compression ratios and a very small reduction in the amplitude distribution. The distribution seems to shift upward, in particular, the low-level sounds below 30 dB. The effective reduction for all compression ratios varies between 1.3 and 1.4, showing no effective reduction of the level distribution, or in other words, no “squashing of the banana”. Figure 2-11 shows a similar analysis of the amplitude distribution of the same speech sample in the octave band around 0.5 kHz with the same legend as the figures 2-9 and 2-10. We see no clear effect of gain. The reduction factors for the compression ratios 2, 4, and 8 are 1.5, 1.9, and 2.0, respectively. This shows that there is some reduction in this frequency band but far less than is given by the compression ratio.

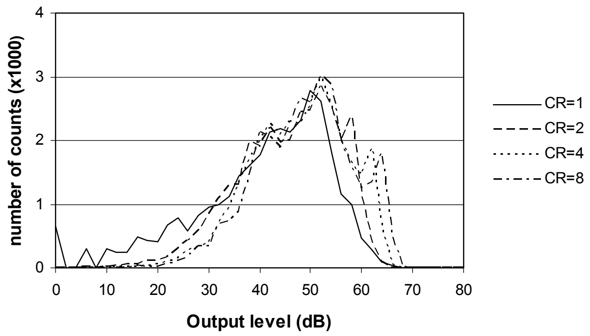


Figure 2-10 *Distribution of speech levels in a 2-kHz octave band for different compression ratios, linear (drawn line), compression ratio 2 (dash), 4 (dot) and 8 (dash-dot).*

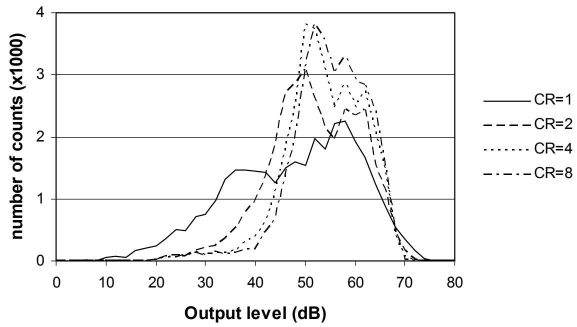


Figure 2-11 *Amplitude distribution in a 0.5-kHz octave band for linear amplification (drawn line) and for compression ratios 2 (dash), 4 (dot) and 8 (dash-dot).*

We conclude that the single-channel, wide-band compressor is effective in reducing level differences. The compression does not lead, however, to a similar reduction of the width of the speech level distribution in the octave bands (no squashing of the speech banana). Theories about fitting the speech banana into the dynamic range are not justified for single-channel compressors.

High-Frequency Control

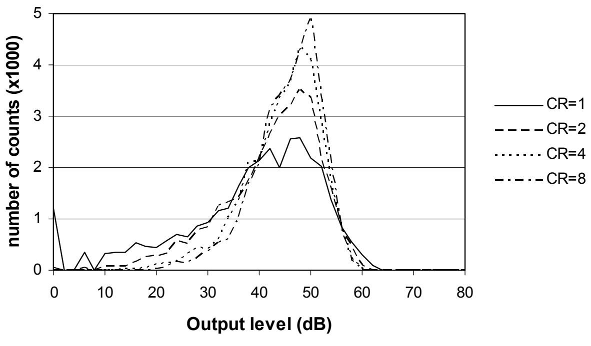


Figure 2-12 *The distribution of speech amplitudes with high-frequency control of the compressor and high-frequency emphasis of the speech, for linear amplification (drawn line) and for compression ratios 2 (dash), 4 (dot) and 8 (dash-dot).*

We also analyzed the speech distribution when the control signal was taken from the output of the FIR-filter in figure 2-2, the filter having a low-frequency cut-off frequency of 1850 Hz. In the linear filter path, a filter was added giving a high-frequency emphasis of 6 dB/oct. In this case the higher frequency components of speech controlled the compression system and not the lower-frequency components. We expected the high-frequency band to be more effectively compressed.

In a way similar to that mentioned above we studied the level distributions of speech in octave bands. The 0.5 kHz octave results showed that there was no change in the amplitude distribution at all. This result was expected because the signal was bypassing the compressor through the linear path. Figure 2-12 shows the 2 kHz octave filtered amplitude distribution of the same speech signal as was used for the analysis in figure 2-9. We see a reduction of the distribution. The effective reduction of the distribution width at the -20 dB level is now 2.0, 2.5, and 3.0, respectively. Also, the reduction in width of the amplitude distribution is now less than was found in the wide-band analysis but more effective than when a wide-band control signal was used.

We conclude that if we want to achieve a squashed speech banana, the signal has to be split up into at least two channels and the compression has to be controlled by a signal with similar spectral contents of the signal to be compressed.

Commercial Hearing Aids

In a similar way, we analyzed the distribution of speech levels of the same speech sample passing through the S45-1 hearing aid and a hearing aid with a K-amp circuit. The distributions were interpreted as above in an effective reduction of the level distribution. In the wide-band analysis, the reduction of the level distribution with the S45-I was 1.5 (static compression ratio is 5) and with the K-amp was 1.9 (static compression ratio is supposed to be just above 2). In the octave band analysis, we found in the 0.5 kHz band a reduction of the distribution width by a factor of 1.3 for the S45-1 and no reduction of the distribution for the K-amp; in the 2 kHz band, the effective reduction of the distribution width was a reduction by a factor of 1.3 for the S45-1 and 1.1 for the K-amp.

The analysis showed that the S45-1 was not reducing the distribution width very effectively, although the static compression ratio was 5. Most probably the time constants are too long for an effective reduction of speech. The K-amp actually behaved as a single-channel syllabic compressor with wide-band control because it effectively reduced the wide-band level distribution but not the distribution in the high-frequency band.

2.6 Discussion

We have seen that for compressors to be effective for speech, the time constants should be very short. We argued that if we want to improve the detection of certain speech features of consonants, amplitude modulations of up to about 30 Hz should be compressed. Our experimental compressor meets this requirement when using an attack time of 5 ms and a release time of 15 ms. However, the effective compression ratio quickly falls at higher modulation frequencies when the release time is decreased. Speech intelligibility data with this compressor show that for low compression ratios (less than 3), the speech score does improve for such fast (phonemic) compressors in comparison to linear amplification with the same frequency response (Verschuure et al., 1993; Verschuure et al.1994, Goedegebure et al.

2001). Analysis of which speech features contributes to the improved score showed that consonant recognition indeed improves because of an improved plosive-fricative distinction (Goedegebure et. al. 2002). For slower compression systems, no improvements in the maximum score can be expected as the dynamics of the speech signal it self is not reduced by the compressor. The advantage of such systems is that the speech score still remains higher over a wider range of presentation levels than it does for linear amplification. The methods described here are able to distinguish between both types of compression systems.

The distribution of compression channels is another important feature of a compression system. The second method based on amplitude levels distributions of speech shows that although a fast *single-channel* compressor is effective in reducing level differences between different phonemes, the compression does not lead to a similar reduction of the width of the speech level distribution in the octave bands (no squashing of the speech banana). This is not surprising if we consider the following example of what a single-channel, wide-band compressor does. Consider a vowel (high-level sound with spectrum falling off at 10 dB/oct over 0.5 kHz) followed by a high-frequency consonant (spectrum around 2 kHz, level -20 dB versus peak spectral level of the vowel). The vowel-consonant combination gives only a small modulation in the 2 kHz band for the uncompressed signal. With wide-band compression the level of the consonant is raised, actually causing more modulation in the high-frequency band with compression than without compression. The example shows that unpredictable effects may occur. Our measurements confirm that no effective compression is obtained within octave bands with such a wide-band single-channel system. Furthermore, we show that it is possible to reduce the width of the level distribution within the high-frequency band by applying compression only within the high-frequency band. Therefore, some kind of multi-channel approach is needed when the aim is to reduce the dynamics of speech throughout the whole spectrum. A single-channel system may still be effective in improving speech intelligibility because certain weak speech sounds are emphasised, but not because it reduces the speech dynamics within frequency bands.

In view of the data on speech intelligibility and the effectiveness of compressors, we would propose to use slow-acting compressors to adapt to different acoustical conditions without the compressor having any effect on the speech signal. Single-channel compression systems have to be considered irrelevant for the reduction of the amplitude distribution of speech per octave band and thus should not be used for that purpose. The only possible positive effect is a better detectability of phonetic features. In that case we have to make sure that the detectability is not confounded by temporal distortions, such as temporal overshoots. To reduce temporal overshoots we used a delay between the control signal and the compression signal. The method with the amplitude-modulated signal can be used to choose an appropriate value of the delay. Figure 2-6 shows that a too long or too short value of the delay reduces the effectiveness of the compressor.

The measurements with the commercial hearing aid systems confirm the potentials of using analytic methods to describe the compression characteristics. The lower effectiveness for the S45=I than for the K-amp system can be expected from the theoretical values of the time constants. However, the dual time constants used in the K-amp system make it difficult to predict the resulting effect on speech. Our measurements show that speech is indeed affected by the system, but not to the same extent as with a real phonemic compressor such as our system. Furthermore, we see that the low effectiveness of the S45-I and the higher effectiveness of the K-amp are reflected in both measurements. This suggests that both methods can be used with some predictive value to characterise a commercial hearing aid system.

There are also some considerations that should be taken in account when using the methods described here. The theoretical calculation of the Appendix shows that a logarithmic compressor distorts the temporal pattern by introducing side bands into the modulation spectrum. This was confirmed by our measurements which made us decide to limit the modulation depth in the analysis. However, speech of a single talker has a modulation depth higher than 45%. This implies that most compressors in hearing aids distort the modulation spectra of speech by introducing extra harmonics. These distortions may influence speech intelligibility in an unpredictable way, requiring the use of speech intelligibility tests when compressors are used. The limitation of the method to 45% modulation reflects the amplitude envelope distortion of logarithmic compressors but is no indication of limited use of the described analysis method.

Another issue is what kind of speech material should be used for the level-distribution method. We used special speech material for the analysis of the speech amplitude distribution. We wanted to use normal continuous discourse. Our first idea was to use a radio interview for that purpose. However, a speech signal recorded from the radio showed unwanted compression effects because the signal passes through many stages of compression: at the recording site, at the input to the transmitter, and in the receiver. The material, therefore, could not be used for our purposes. On the other hand, we did not want to make a recording ourselves because such recordings often do not have the characteristics of natural speech, as they are read. We obtained a recording on DAT-tape from a radio engineer working at the Dutch Broadcasting Enterprises. He assured us that the recordings had been made without any compressors being used. The amplitude distribution showed a width of about 45 dB, and this width was wider than the width of speech often reported in the literature. The difference was due to the amount of intonation (Verschuure & Dreschler, 1993). We also used another sample with more hesitations in it and found that the distribution of the levels was very similar except that the number of counts of intervals with a very low level (below 20 dB in the Figs. 12-16) increased dramatically. The width and shape of the level distributions above 25 dB were very similar. The use of more artificial speech(-like) material, such as the ICRA-noises (Dreschler et al. 2001), may help to standardise the method .

2.7 Conclusion

A technique with an amplitude-modulated signal can be used to assess the effectiveness of compressing systems at a certain level. The modulation depth should not exceed 45% for an accurate analysis.
The effect of compression on the dynamic range of speech can be assessed by analysis of the amplitude distributions within octave bands. Although this method has a more qualitative approach, it has the advantage that it takes the frequency-dependent characteristics of speech and the compressor in consideration. Both methods show that phonemic compression can effectively reduce amplitude-modulations and speech-level differences within octave bands as long as the time constants are fast enough and a multi-channel design is used.

2.8 Acknowledgements

Many people have contributed to this study. It is the result of years of discussions among numerous people on the issue of effectively describing compression. We want to thank explicitly B. Geerdink, M.Sc and P. Termeer, B.Se from Philips Hearing Instruments.
This research has been supported by the Innovatief Onderzoeks Programma Hulpmiddelen Gehandicapten (IOP-HG) and Stichting Innovatief Programma Technologie (STIPT), both programs financed by the Dutch Ministry of Economical Affairs, the Heinsius Houbolt Trust Fund, and the Technological Innovative programme for the Disabled and Elderly (TIDE) of the European Unjon in cooperation with the hearing aid departments of Phihps in Eindhoven, the Netherlands and Siemens in Erlangen. Germany.

2.9 References

Bustamante D. K, & Braida L. D. (1987). Multiband compression limiting for hearing-impaired listeners. J Rehab Res Dev 24:149-160.

Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. (2001) ICRA noises: artificial noise signals with speech-like spectral temporal properties fro hearing instrument assessment. International Collegium for Rehabilitative Audiology. Audiology 40(3): 148-157.

Gennaro S. V. de, Braida L. D., & Durlach N. 1. (1986). Multichannel syllabic compression for severely impaired listeners. J Rehab Res Dev 23: 17-24.

Gennaro S. V. de, Krieg K. R., Braida L. D., & Durlach N. 1. (1981). Third-octave analysis of multichannel amplitude compressed speech. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing: 125-128.

Goedegebure A, Hulshof M, Verschuure J, Dreschler WA. (2001) Effects of Single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners. Audiology 40(2):10-25.

Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA. (2002) The effects of phonemic compression and anti-USOM on the perception of articulatory features in hearing-impaired listeners. Int J Audiol 41(7): 414-428.

Killion M. (1993). The K-Amp hearing aid: An attempt to present high fidelity for persons with impaired hearing. Am J Audiol 2:52-74.

Maré M. J., Dreschler W. A., & Verschuure H. (1992). The effects of input-output configuration in syllabic compression on speech perception. J Speech Hear Res 35: 675-685.

Pavlovic C. V. (1992). Statistical distribution of speech for various languages. J Acoust Soc Am 88 (Suppl. l): 8SP10.

Plomp R. (1983). Perception of speech as a modulated signal. Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht :29-40.

Steeneken H. J. M., & Houtgast T. (1980). A physical method for measuring speech transmission quality. J Acoust Soc Am 67: 318-326.

Stone A. M., & Moore B. C. J. (1992). Syllabic compression: Effective compression ratios for signals modulated at different rates. Brit J Audiol 26: 351-361.

Verschuure J., & Dreschler W. A. (1993). Present and future technology in hearing aids. Journal of Speech-Language Pathology and Audiology IRevue d' Orthophonie et d'Audiologie, JSPLA Monograph(Suppl. l): 65-73.

Verschuure J., Dreschler W. A., Haan E. H. de, Cappellen M. van, Hammerschlag R., Maré M. J., Maas A. J. J., & Hijmans A. C. (1993). Syllabic compression and speech intelligibility in hearing impaired listeners. Scand Audiol 22 (Suppl. 38): 92-100.

Verschuure J., Maas A. J. J., Stikvoort E., & Dreschler W. A. (1992). Syllabic compression in hearing aids: Technical verification of nonlinear signal processing. Acustica 7414: s.14.

Verschuure J., Prinzen T. T., Dreschler W. A. (1994). The effects of syllabic compression and frequency shaping on the speech intelligibility in hearing impaired people. Ear Hear 15:13-21.

2.10 Appendix

Theory of Spectral Modulation Analysis

The effectiveness of a compressor can be tested by using a sine wave with carrier frequency (ω_c) and a modulating frequency (ω_m). The amplitude of the carrier is a_0, and the amount of modulation is given by the modulation index m. The amount of modulation can be determined accurately by analyzing the spectrum of the input or output signal.
An amplitude modulated sine wave can be described by:

$$x(t) = a_0(1 + m\cos(\omega_m t))\cos(\omega_c t) \tag{2.4}$$

The spectrum of the signal of Equation 4 consists of three frequency components:

$$\begin{gathered} \frac{1}{2} m\, a_0 \cos((\omega_c - \omega_m)t) \\ a_0 \cos(\omega_m t) \\ \frac{1}{2} m\, a_0 \cos((\omega_c + \omega_m)t) \end{gathered} \tag{2.5}$$

We see that the difference in amplitude between the two symmetrical side bands and the carrier can be translated in dB. The equation for the difference in level in dB is:

$$\Delta S = 20 \ \log m - 6 \tag{2.6}$$

The comparison of the spectral level of the central component with that of one of the side bands provides a direct measure of the modulation depth.

Let us now consider how a sinusoidally-modulated sine wave (Eq. 2.4) is processed by a compressor. Compressors usually work in the logarithmic domain (dB versus dB) and not in the amplitude domain. We distinguish between linear amplitude measures denoted by small italics and logarithmic amplitude measures denoted by capital italics. The output signal (*U* in dB) depends on the input signal above the compression threshold (I_k) according to:

$$U = CR \ (\ I - I_k) + I_k \tag{2.7}$$

where *I* is the input signal and *CR* is the compression ratio. Transforming Equation 2.7 to the linear amplitude domain, gives:

$$20 \log \frac{u}{u_0} = CR \left(20 \log \frac{i}{i_0} - 20 \log \frac{i_k}{i_0} \right) + 20 \log \frac{i_k}{i_0} \tag{2.8}$$

Equation 2.8 can be rewritten in linear amplitudes as:

$$\frac{u}{u_0} = \frac{i^{CR}}{i_0^{CR}} \left(\frac{i_k}{i_0} \right)^{1\text{-}CR} \tag{2.9}$$

Substituting the time signal *x(t)* (Eq. 2.4) for i and assuming that the signal is well above the compression threshold reduces i_o and i_k to amplitude factors. We get:

$$u = u_o \left(\frac{i_k}{i_0} \right)^{1\text{-}CR} \left(\frac{a_0}{i_o} \right)^{CR} (\ 1 + m \cos \omega_m t \ )^{CR} \cos^{CR} \ \omega_c t \tag{2.10}$$

The first part of Equation 2.10 is a simple amplitude factor describing the action of the compressor. Substituting a new compressed amplitude, b_o, gives:

The second part of Equation 2.10 or 2.11 shows that both the carrier sine wave and the modulation sine wave can be distorted by the compressor. This effect can be understood easily:

$$u = b_o \ (\ 1 + m \cos \omega_m t \ )^{CR} \cos^{CR} \ \omega_c \ t \tag{2.11}$$

1. If we have a fast compressor and a low-frequency carrier, and for a sufficiently large compression ratio, the compressor acts instantaneously on the amplitude of the carrier and no sine wave will appear at the output. In our case, we assume that the carrier frequencies, which are above 100 Hz, are much higher than the upper limit of the effective compression frequency, which is below 50 Hz. For this reason, the effective compression ratio is 1 and can be omitted from that part of the equation.
2. If we use a high modulation depth and sufficiently fast compression, the modulation waveform at the output will no longer be a sinusoid because of the nonlinearity of the logarithmic (dB) scale. As a consequence the higher harmonics of the modulation frequency appear, and the modulation waveform will be distorted.

The modulation factor can be analyzed using a Taylor expansion. The general form of the Taylor expansion is:

$$(1+x)^{\alpha} = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \dots \qquad (2.12)$$

This equation may only be used if x is small compared with 1. In our application, this implies the use of small modulation depths. Computing the first three terms of the Taylor approximation, replacing x with $m\cos\omega t$ and α with CR and using geometric equations for the second and third powers of the cosine, we get:

$$\begin{aligned}\frac{u}{b_0} &= \left(1+\frac{1}{4}CR(CR-1)m^2\right)+\\ &+\left(CRm+\frac{1}{8}CR(CR-1)(CR-2)m^3\right)\cos\omega_m t+\\ &+\frac{1}{4}CR(CR-1)\ m^2\cos 2\omega_m t+\\ &+\frac{1}{24}CR(CR-1)(CR-2)m^3\cos 3\omega_m t+\dots\end{aligned} \qquad (2.13)$$

This equation gives the amplitude factors of each modulation side band. We also see that there is an additional DC-term.

We now can make error estimates of the additional terms, specifically those adding to the DC-component and the first side band.

Because CR has a value between 0 and 1, we know that:

$$CR(CR-1) \le 0.25 \qquad (2.14)$$

and

$$CR\,(\,CR-1\,)(\,CR-2\,) \leq 0.75 \tag{2.15}$$

If we do not want the errors to become larger than 5% in amplitude, we have to limit the modulation depth to less than 45% (m < 0.45). For this value of m, the harmonics of the modulation side bands are more than 40 dB down for the second harmonic and more than 50 dB down for the third harmonic.

The calculation shows that the logarithmic compressor causes distortion in the modulation amplitude envelope. The spectral analysis method gives a correct representation of its effectiveness as long as the modulation depth of the input signal is rather small (< 45%). This implies that the high modulation depths of speech cannot be imitated by the signal because of the high amount of distortion. It does not rule out the use of the method for higher modulation depths because, in general, we are interested in the effectiveness of a compressor near a certain level. The range of levels that make up speech can be analyzed by determining the effective compression ratios at a number of levels within the speech range. This also implies that speech compressed by a logarithmic compressor shows temporal distortion.

Using modulation depths of less than 45%, we can determine the effectiveness of the compressor at that particular level directly from the spectrum. Equation 2.13 simplifies to:

$$\frac{u}{b_0} = (\,1 + CR\,m\,\cos\omega_m t\,)\,\cos\omega_c t \tag{2.16}$$

In spectral terms this leads to side bands with a level of 0.5 $CR\ m$. Writing the effective compression ratio CR_{eff} explicitly using Equation 2.6, we get:

$$CR_{eff} = 10^{\frac{\Delta S_o - \Delta S_i}{20}} \tag{2.17}$$

Within the limitation of small modulation depths (m < 0.45), we can now analyze the effectiveness of the compressor system simply by looking at the input and output spectrum. The determining factor is the level differences between the carrier and the first upper or lower side band. This method allows for an easy determination of the effectiveness of a compressor at a certain level for a given carrier frequency and for a given modulation frequency.

The temporal distortion of compressors appears as higher harmonies, and these harmonics could be used as a measure of total temporal distortion.

3 The effect of compression on speech modulations

Submitted for publication in Int. J. Audiology
Goedegebure A, Verschuure J, Dreschler WA

Abstract

Although dynamic range compression is frequently used in modern hearing aids, no good methods are available to measure the effect of compression on speech signals. We have developed an analytic method to estimate the amount of compression for speech or speech-like signals. Within separate frequency bands an effective compression ratio is calculated as function of modulation frequency. The method has been tested using an experimental fast-acting compression system. The results show that the relevant modulations in speech are affected only by compression with relatively fast time constants. The results depend on the stimulus used. Both speech and artificially modulated noise (ICRA) can be used. It can be concluded that the method has potentials to characterise the compression used in hearing aids for speech(-like) signals.

3.1 Introduction

Dynamic range compression is a commonly used signal-processing technique in modern hearing aids. The introduction of digital techniques has increased the diversity and complexity of compression systems. A wide range of possible compression configurations is available, each with its own specific features. The number of compression parameters is large and the exact implementation of the system is often not specified by the manufacturers of hearing aids.

The amount of compression is usually defined by the "compression ratio". This value defines the relation between the input- and output dynamic range of the system within the compression range. For static acoustical signals like a continuous sine wave the compression ratio is well defined and can accurately be measured. The resulting amount of compression for *non-static* signals highly depends of the temporal behaviour of the compressor, usually defined by the two conventional parameters "attack time" (reaction time of the compressor to a sudden rise in input signal level) and "release time" (reaction time to a sudden drop in input level). Another important characteristic of a compression system is the number of

independent frequency channels in which compression in performed. A “single-channel” compressor does only include one frequency channel, which often consists of the complete broad-band signal. A “multi-channel” compressor divides the signal in a number of band-filtered parts with a number of (independent) compression systems active in each part of the frequency spectrum.
It is remarkable that many studies address the question whether compression improves or disturbs speech understanding, but that nevertheless only little attention is given to the resulting effect of compression on the speech signal it self. As far as compression parameters are concerned, standard hearing-aid measurements with a 2-cc coupler equipment includes measurement of the static compression ratio, the attack time and the release time. A disadvantage of these methods is that rather simple and artificial signals are used, while the most important signals in real environments have a more complex dynamic and spectral character. The most important signal, speech, is a very complex signal with a specific temporal behaviour over a broad range of frequencies. For such a complex broadband non-static signal it is difficult to predict the combined effect of the different compression parameters on the output signal. The interaction between the different compression parameters can only be measured with an analytic method using a signal with the same spectral and dynamic behaviour as speech.

Only a few analytic methods have been described to measure the effect of nonlinear processing on speech. Payton et al. 2002 and Drullman et al. 1994 describe a method for a speech stimulus using the Speech Transmission Index (STI). Dreschler et al. 2001 used artificial broadband noise signals (ICRA) to measure the effects of noise suppression algorithms in hearing aids. Dyrland et al. 1994 used both speech and noise to analyse the response of non-linear hearing aids to broad-band signals. All these methods give insight in the dynamic properties of the signal processing but do not come with a usable quantitative measure to express the effect of compression on speech.
Verschuure et al.(1996) describe two methods to determine the effect of compression on speech. The first method uses Amplitude Modulated (AM) -sinusoids to determine the effective compression ratio as function of modulation frequency (see also Stone and Moore 1992). The concept behind this method is that speech can be described as the sum of amplitude-modulated signals in different frequency bands (Steeneken and Houtgast 1980, Plomp 1983). The advantage of the described method is that a very accurate quantitative estimation is obtained from the dynamic behaviour of the system within a certain frequency band. However, it can not predict the dynamic behaviour for a complex broad-band modulating signal like speech. Therefore, a second method was introduced by Verschuure et al. (1996) based on the determination of the amplitude distributions of running speech. This method showed that the width the of level distributions within a specific frequency band decreases if compression is effective in that frequency band. The advantage is that it identifies the frequency-dependent response of the compression system to the broadband speech signal. However, the obtained information is more difficult to quantify and does not distinguish between different modulation frequencies.

Obviously, there is a need for a more accurate objective measure to define the effect of compression on speech. The goal of the present study is to find such a measure using an analytic method with speech or speech-like signals as input signal. The proposed method integrates the advantages of aforementioned methods as it quantifies the effect of compression on speech modulations as a function of modulation frequency and frequency band. The research questions are:

- can a method be found to accurately quantify the effect of compression on speech modulations
- what stimulus conditions are needed to optimise the results obtained by the method
- what are the clinical implications of the results obtained by the method

3.2 Methods

Rationale

Speech can be considered as a stream of sounds with a continuously varying spectrum. These spectral differences lead to fluctuations of the envelope of the signal within individual frequency bands. Based on this concept, Steeneken and Houtgast (1980) derived the Speech Transmission Index (STI), a widely accepted and powerful method to predict the effect of room acoustics on speech intelligibility. The success of this method shows the value of quantifying a speech signal in terms of envelope modulations. Plomp (1983) derived so-called modulation spectra of speech by determining the spectral contents of the *envelope* signal within individual frequency bands. The spectra show that that the relevant modulation frequencies of speech are roughly in a range between 0.1 to 40 Hz and that the strongest fluctuations are between 3 to 5 Hz. The shape of the spectra is about the same for all octave bands, but the amount of modulations differs between the bands. The strongest modulations are found within the 1 kHz band while the low-frequency bands contain less modulation.

The use of dynamic range compression will reduce the amount of speech modulation if the compression acts fast enough. For amplitude-modulated sine-waves, Verschuure et al.(1996) showed that for modulation depths smaller than 0.45 the relation between the amount of compression and the modulation function at the output of the compressor can be approximated by (their equation 16):

$$y_{out}(t) = a_o * (1 + CRm\cos(2\pi F_m t))\cos(2\pi f_c t) \tag{3.1}$$

$y_{out}(t)$ is the amplitude at the output of the compressor, a_o the static amplitude that depends on the average gain applied to the signal, F_m the modulation frequency, f_c the carrier frequency, CR the compression ratio ands m the amplitude modulation depth at the input. Based on this relation, Verschuure et al (1996) defined a measure for the actual amount of compression applied to an amplitude-modulated sine wave:

$$CR_{eff} = 10^{\frac{\Delta S_{in} - \Delta S_{out}}{20}} \tag{3.2}$$

The effective compression ratio CR_{eff} can be derived from the measured amplitude difference in dB of carrier and first side-band at the input (ΔS_{in}) and at the output (ΔS_{out}) of the system. The loss of modulation amplitude caused by the compressor is used to describe how effective the compressor is as function of modulation frequency. For low modulation frequencies and short time constants the effective compression ratio approaches the static compression ratio as defined by the compression tables. By increasing the modulation frequency or increasing the time constants of the compressor the effective compression ratio drops to values just around one.

In the present study we have used a similar approach to calculate the effective compression ratio for speech-like signals. From the modulation spectra of speech we have estimated the reduction in modulation amplitude resulting from compression (for details see the next section). The difference in modulation-spectrum amplitude level with and without compression, ΔS_{Fm} , can be translated to an effective compression ratio like in formula 3.2:

$$CR_{eff} = 10^{\frac{\Delta S_{F_m}}{20}} \tag{3.3}$$

This measure estimates the effectiveness of the compressor as function of modulation frequency. It can be applied to any desired input signal, whereas the use of equation 3.2 is limited to “simple” amplitude-modulated signals only. This means that we can quantify the effect of compression on more relevant and complex signals such as speech. Furthermore, the effect is quantified as function of modulation frequency, which was not the case for a previous method based on energy distributions of speech (Verschuure et al. 1996). This allows us to analyse the time-dependent behaviour of the compressor for speech signals. A restriction of the method may be that significant non-linear harmonics will show up at high modulation depths because the input-output relationship is defined in dB’s. Verschuure et al. 1996 showed that a linear approach like in equaton 3.2 can only be used for modulation factors less than 0.45 (-6.9 dB) without introducing non-linear distortions of more than 5%. This effect will be investigated by using various kinds of stimuli.
A second measure, derived from the CR_{eff} spectra, is introduced to quantify the time-dependent behaviour of the compressor. We expect to find a decreasing value of CR_{eff} for increasing modulation frequencies. Therefore a cut-off modulation frequency, *F_m cut-off*, is defined based on the *relative* effectiveness:

$$CR_{eff\%} = 100 \frac{CR_{eff} - 1}{\max(CR_{eff}) - 1} \tag{3.4}$$

This measure has a maximum of 100% for the modulation frequency at which CR_{eff} reaches the maximum level ($\max(CR_{eff})$) and a value of 0% for the frequencies the compressor is not effective at all ($CR_{eff} = 1$). F_m *cut-off* is defined as the first modulation frequency at which the effectiveness drops below 75%. This means that F_m *cut-off* corresponds to a 25% *decrease* in effectiveness between the maximum value of CR_{eff} and 1.

Implementation of method

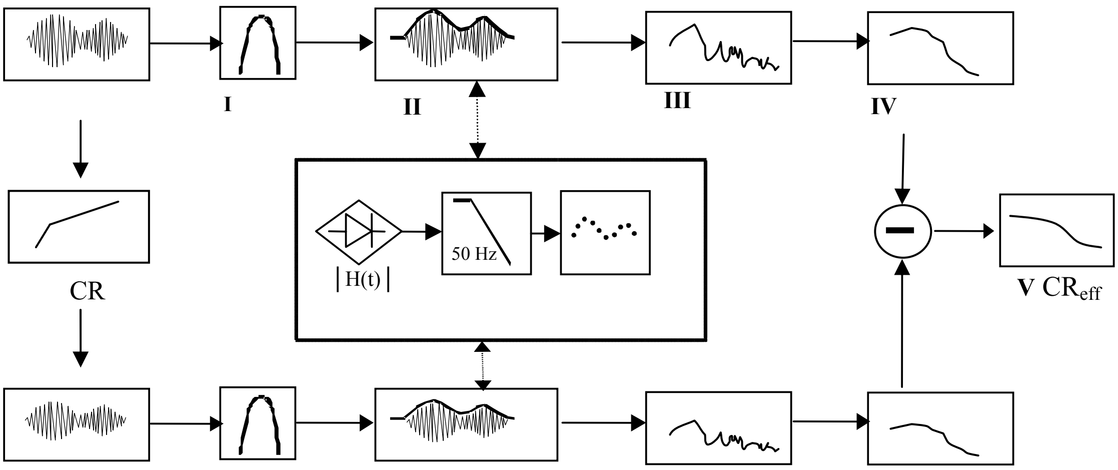


Figure 3-1 *Block scheme of the procedure to calculate the effective compression ratio from the envelope spectrum of speech*

We implemented a MATLAB program (The Math Works Inc., version 5 release 11) to calculate the envelope spectrum and the effective compression ratio as function of modulation frequency. The block diagram of this procedure is shown in figure 3-1. The same analysis is applied to both the unprocessed and the compressed speech signal (on top en below respectively).

I. The speech is band-pass filtered using a 5th order elliptical Butterworth filter with adjustable centre frequency and bandwidth. The band-filter can also be bypassed resulting in a broad-band analysis.

II. For each frequency band the envelope is estimated by taking the magnitude of a standard Hilbert-transform. The Hilbert envelope is low-pass filtered with a 50-Hz elliptical low-pass filter and then downsampled to a frequency of 200 Hz. The low-pass filtering removes any fine-structure components of the speech signal, such as the fundamental frequency of the speaker. The resulting function will be referred to as “the envelope function”.

III. The Power Spectral Density (PSD) of the envelope function is estimated using a standard Welch procedure (Signal-Processing toolbox, Matlab 5.11 The Math Works Inc.). This procedure divides the function in overlapping time windows of which the power density is estimated using a discrete Fourier transform. After that, the spectra of each of the windows are averaged. The advantage of taking the average spectrum of windowed parts instead of determining one spectrum of the complete signal is that the variance of the spectrum decreases (Priestly 1981). The parameters used for the PSD procedure are a window length of 1600 points (8 seconds), a Hanning window as window type with 40 % overlap of the windows. This parameter setting gave relatively smooth spectra with a frequency resolution of 0.125 Hz.

IV. The intensity values of the PSD are summed over modulation frequencies for each octave band and normalised using the 0-component of the PSD. The normalisation is set to reach a value of 0 dB (m=1) for an amplitude modulated sine wave. This calculation results in a modulation spectrum defined in the intensity domain, just like in figure 3-1. However, dynamic-range compression and the equations for the effective compression ratio are defined in the amplitude domain. Therefore the spectra are transformed into the amplitude domain, indicated as amplitude modulation spectrum.

V. Now equation 3.3 can be applied by taking the difference between the spectra from the unprocessed and the compressed signal. The result is an estimate of the effective compression ratio (CR_{eff}) as function of modulation frequency for a given frequency band. Additionally a cut-off modulation frequency is derived using formula 3.4.

Stimuli

A first stimulus contains a 1-minute lasting sequence of 34 test sentences of a recently developed Dutch speech-reception threshold test (VU98 sentences, Versfeld et al. 2000) without time gaps in between the sentences. A second stimulus consists of the same speech stimulus but now mixed with matched speech-shaped stationary noise at a signal-to-noise ratio (SNR) of + 6 dB. The two samples are indicated by “Speech” and “Speech SN6” respectively. A second type of signal is the artificial speech-like modulated noises of the ICRA-CD (Dreschler et al. 2001). This signal has the advantage that it has a standardised long-term speech spectrum. We have used the first 1 minute of the simulated single-talker female voice (track 4, “ICRA 1sp”), the two-talker noise female and male (track 6, “ICRA 2sp”) and the 6-talker noise (track 7, “ICRA 6sp”) respectively. This last sample consists of 3 simulated male and 3 simulated female talkers of which 2 at a level of -6 dB relative to the first talker. In addition we created a speech sample by adding matched speech-shaped stationary noise to the ICRA 1sp sample at a SNR of +6 dB. This sample is called ICRA 1spSN6.

Table 3-1 *Characteristics of stimuli used*

Name	*Characteristic*	*number of speakers*	*noise*	*max. value m (dB)*
Speech	speech sentences	1	-	-1.9
Speech SN6	speech sentences	1	SNR=6 dB	-6.3
ICRA 1sp	modulated noise	1	-	-1.3
ICRA 1spSN6	modulated noise	1	SNR=6 dB	-5.9
ICRA 2sp	modulated noise	2	-	-4.3
ICRA 6sp	modulated noise	6	-	-8.1

A description of the used stimuli is summarised in table 3-1. Figures 3-2a/b and 3-3a/b show the amplitude modulation spectra of the speech stimuli and two of the ICRA stimuli respectively. Each of the figures shows the spectra for the individual octave bands at 0.5, 1, 2 and 4 kHz. The spectra of all signals have a very similar shape. They all show a maximum modulation depth near 4 Hz. This is in agreement with the modulation spectra shown by

Plomp (1983). In general the 1-kHz band contains the highest amount of modulations (about 0 dB for 1 speaker without stationary noise), whereas the 500-Hz band contains the least amount of modulations. As expected the modulation depth decreases by adding noise to the signal (figure 3-2b) or by increasing the number of speakers (figure 3-3b). The maximum values of the modulation depth within the 1-kHz band are listed in table 3-1. An SNR of +6 dB results in maximum modulations of about -6 dB, which is the range at which non-linear distortion is reduced to 5% or less with compression (m<0.45, see first section in methods). A similar effect is obtained by mixing various speakers (ICRA 2sp and ICRA 6sp). Note that the general shape of the spectra is hardly changed although the amount of modulation reduces by adding noise or speakers.

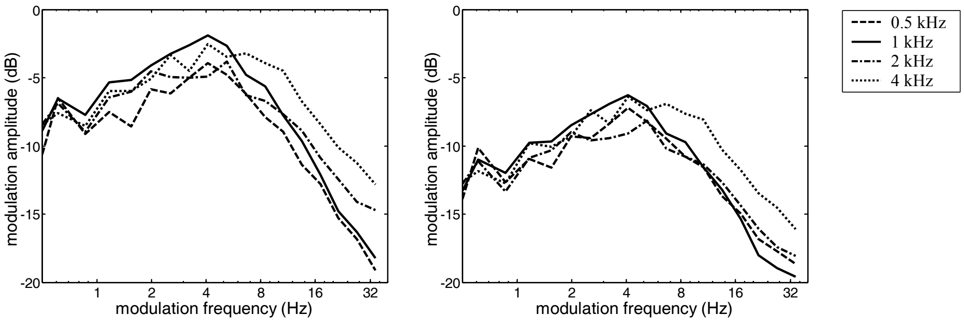


Figure 3-2a/b *Amplitude modulation spectra for the Speech stimulus (a) and the Speech in noise stimulus "Speech SN6" (b) as obtained in frequency octave-bands 0.5, 1, 2 and 4 kHz.*

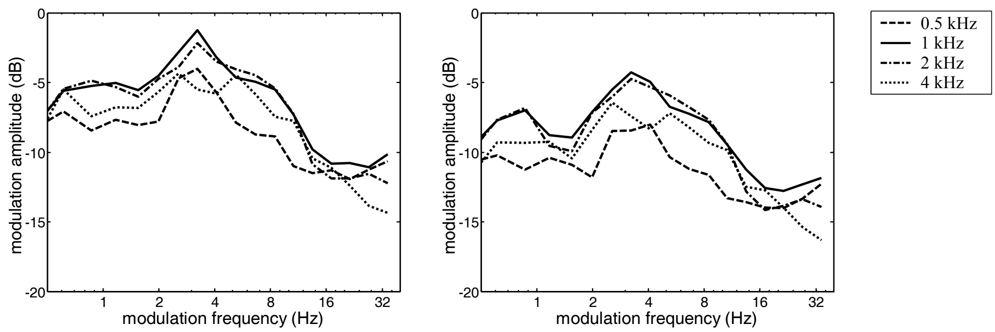


Figure 3-3a/b *Amplitude modulation spectra for the ICRA 1sp stimulus (a) and the ICRA 2sp stimulus (b) as obtained in frequency octave-bands 0.5, 1, 2 and 4 kHz.*

Compressor

The stimuli are processed by an experimental digital hearing-aid system designed by Rass et al. (2000). The system consists of a programmable DSP-platform with 24-bit AD-conversion and a Motorola DSP56003 chip with a fixed 24-bit internal architecture (DASi-2 system). The DSP contains a three-channel AGC algorithm, implemented in the FFT-domain (Rass et al. 2000). The essential compression parameters as compression ratio, time constants, compression threshold and cross-over frequencies can be varied. The time constants are defined according to the standard description for hearing aids (ANSI S3.22, 1996). The default parameter settings are CR=2, attack time = 1 ms, release time = 15 ms and compression threshold at 30 dB SPL ensuring fast-acting wide dynamic range compression (WDRC).

The compressors within the three compression channels act independently. Various compression-channel settings are used. A first configuration consists of cross-over frequencies 703 Hz and 1359 Hz with active compression within the mid-frequency channel *only*. Although such a setting is hardly used in practise it enables analysis within a 1-kHz octave band without any possible interaction between various compression bands. The results for this configuration will be presented in sections A and B. Three-channel compression is applied in section C, using a configuration with the same cross-over frequencies as before (703 Hz and 1359 Hz) and a second configuration with cross-over frequencies similar to that of the ICRA modulation-bands (890 Hz and 2290 Hz compared to 850 Hz and 2500 Hz respectively). To approach a single-channel broad-band configuration both cross-over frequencies are set to the highest values possible (6047 Hz and 8391 Hz) and the compression is only applied within the "low-frequency" channel.

Experimental set-up

The speech material is played from audio CD and directly fed into the experimental compressor using the line-out stereo channel. The input rms-level is set to an equivalent level of 70 dB SPL. The output of the compressor is recorded digitally by a PC at 44 kHz and 16 bit (AC '97, Digital Audio). Note that the sound is *not* transformed into an acoustical signal in any of the stages. Only the electronic signal is used. From the recorded and the original signals the effective compression-ratio is calculated off-line using the MATLAB program described above.

3.3 Results

A. Effect of compression ratio and compression kneepoint within a 1-kHz frequency band

Figures 3-4a and 3-4c show the effect of compression on the modulation spectra within the 1-kHz band for the stimuli Speech and Speech SN6 respectively. The corresponding calculated values of CR_{eff} using formula 3 are shown in figures 3-4b and 3-4d respectively. The modulation spectrum shifts downwards as a result of compression, which shows that the compressor reduces the degree of modulations. For the Speech alone stimulus we find a smaller shift in modulation spectrum than for Speech in noise (SN6), resulting in considerably lower values of the effective compression ratio. Instead of the static values of 2 and 4 the maximum values are just above 1.5 and 2.5 respectively. For Speech in noise (SN6) the corresponding values of the effective ratio are just above the theoretical values as long as the modulation frequency is relatively low (< 4 Hz). At higher modulation frequencies the effectiveness decreases, most probably because the compressor has difficulties to follow the faster modulations. Interestingly, for Speech alone the drop off of CR_{eff} starts for higher modulation frequencies than for Speech in noise. Inspection of the modulation spectra in figures 3-4a and 3-4c helps us to understand why this happens. Both figures 3-4a and 3-4c contain the modulation spectrum corresponding to a stationary speech shaped noise. This line can be considered as a kind of noise floor for the speech modulations. The spectra of the compressed Speech SN6 stimuli are closer to the noise level as the Speech spectra, resulting in a less effective compression at high modulation frequencies.

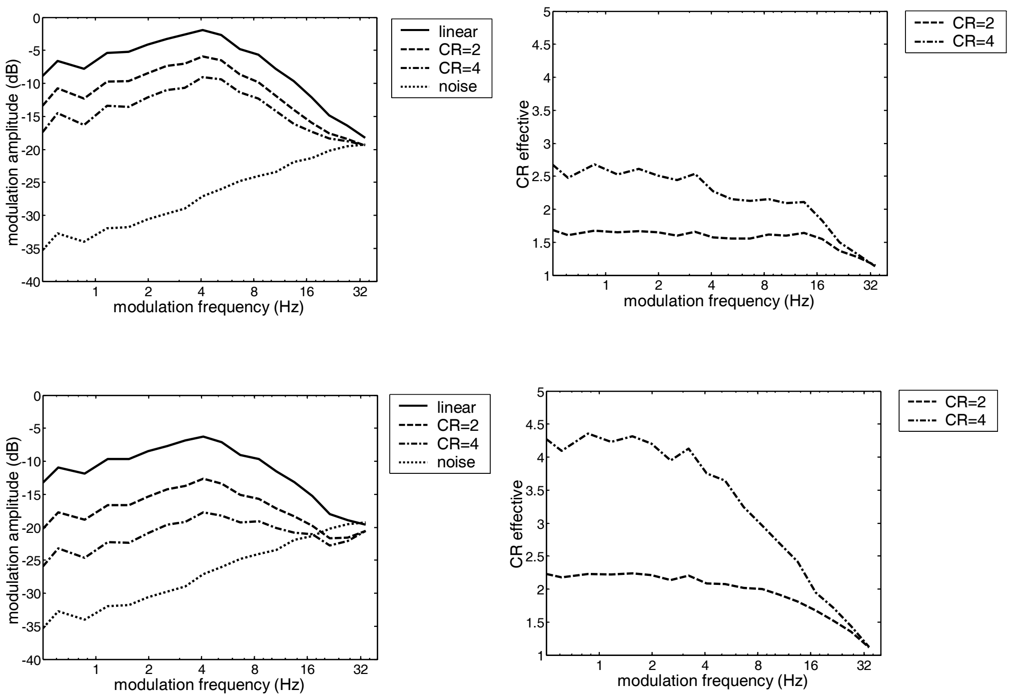


Figure 3-4a/b/c/d *Amplitude modulation spectra for Speech alone (a) and Speech in noise (SN6) stimulus (c) within a 1-kHz octave band with and without compression. The corresponding values for the effective compression ratio CR_{eff} are shown in figures 3.4b (Speech) and 3.4d (Speech SN6). In all cases an attack time of 1 ms and a release time of 15 ms has been used.*

The results for the ICRA stimuli are summarised in figures 3-5a (CR=2) and 3-5b (CR=4) in a similar way as in figures 3-4b and 3-4d. It is remarkable that there is not much difference between the ICRA 1sp, ICRA 2sp and ICRA 6sp stimuli. For these three stimuli the effective compression ratio is systematically lower than the value as defined by the static CR.

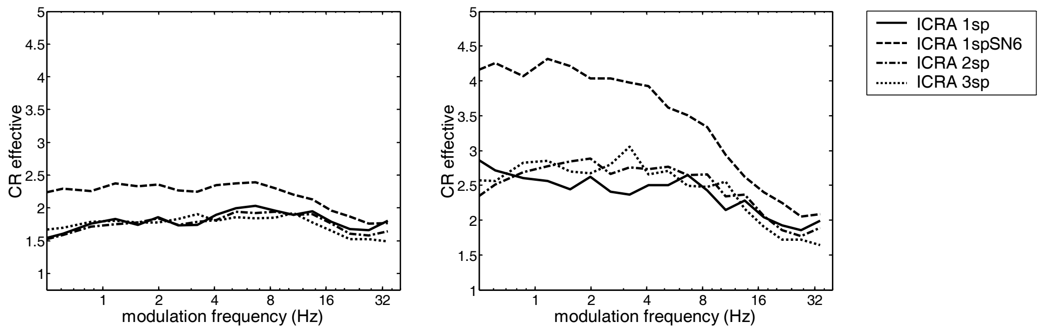


Figure 3-5a/b *Effective compression ratio as function of modulation frequency for the ICRA stimuli, using compression with CR=2(a) and CR=4 (b) within a 1-kHz octave band.*

Increasing the number of speakers does not result in a significantly different behaviour of the compressor. On the other hand, adding noise to the 1-speaker condition does have a considerable effect on the effective compression ratio. For the ICRA 1spSN6 stimulus higher values of CR_{eff} are obtained. The effective compression ratio is even higher than the theoretical value. Note that there is a remarkable similarity between the results for the ICRA 1spSN6 and the Speech SN6 stimuli on one hand, and between the results for the ICRA 1 sp and the Speech alone stimuli on the other hand.

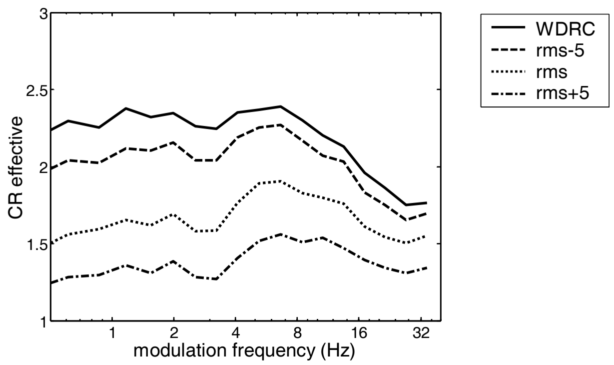


Figure 3-6 *Effect of kneepoint of the compressor on the effective compression, using CR=2 within a 1-kHz octave band with attack time of 1 ms and release time of 15 ms. The stimulus is ICRA 1spSN6.*

Figure 3-6 illustrates the effect of varying the compression kneepoint for the ICRA 1spSN6 stimulus. The kneepoint is set to respectively 30 dB SPL ensuring Wide Dynamic Range (WDRC), rms-level of stimulus -5 dB, rms-level and rms-level+5 dB. The figure clearly shows that the smaller the dynamic range, the less effective speech modulations are compressed. The effectiveness equally decreases for all modulation frequencies. Similar results are found for the other stimuli.

B. Temporal behaviour of compressor within 1-kHz band

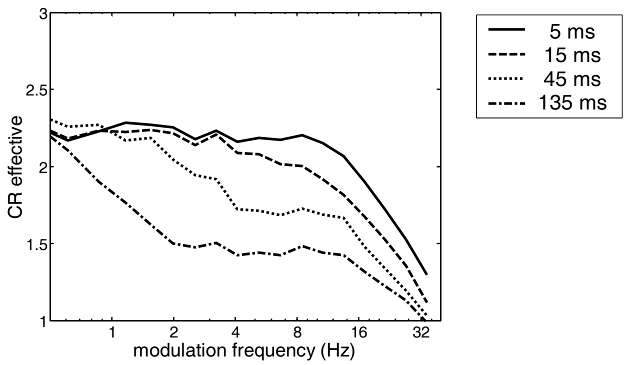


Figure 3-7 *Effect of release time on the effective compression ratio for the Speech SN6 stimulus, using compression with CR=2 and a fixed attack time (1 ms) within a 1-kHz octave band. Note that the dashed curve, corresponding to a release time of 15 ms, is the same curve as presented in figure 3.4d.*

The temporal behaviour of the compressor is one of the most important characteristics to be investigated with the present method. The settings of the time constants mainly determine how effective speech modulations are compressed. Figure 3-7 shows the effect of changing the release time on CR_{eff} for the stimulus Speech SN6. As expected, the effectiveness of the compressor clearly reduces for higher values of the release time.

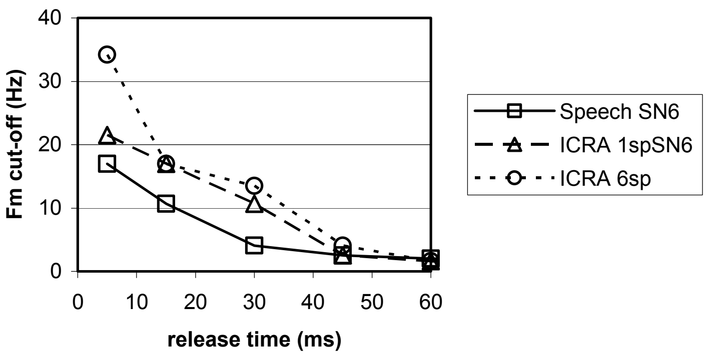


Figure 3-8 *75%- cut-off modulation frequency for CR_{eff} as function of release time of the compressor, using CR=2 within a 1-kHz octave band and a fixed attack time of 1 ms. Differences are shown between the stimuli Speech SN6, ICRA 1spSN6 and ICRA 6sp.*

To compare the time-dependent behaviour of the compressor for different stimuli we calculated a cut-off modulation frequency (F_m cut-off), above which the value of CR_{eff} significantly decreases (see methods, formula 4). Figure 3-8 shows the values of F_m cut-off as function of release time for the stimuli Speech SN6, ICRA 1spSN6 and ICRA 6sp. For the ICRA 1sp and 2sp similar results are found as for the ICRA 6sp stimulus, but they are not shown here. The effective F_m-range of the compressor quickly decreases with increasing release time for all stimuli. A value of 30 ms or lower is needed to effectively compress the most dominant speech modulations. In general a slightly broader F_m-range is found for the ICRA stimuli than for Speech, indicating that a relatively more effective compression is found at higher modulation frequencies.

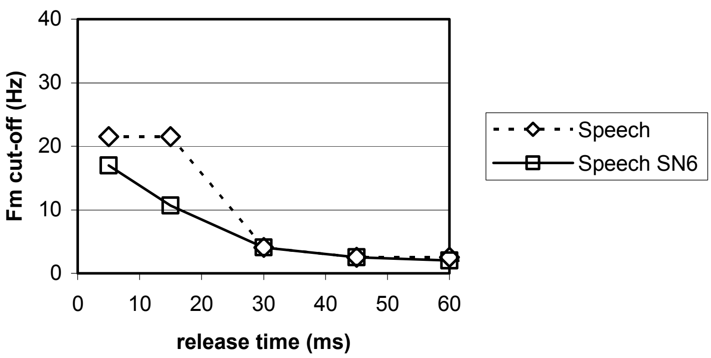


Figure 3-9 *75%- cut-off modulation frequency for CR_{eff} as function of release time of the compressor, using CR=2 within a 1-kHz octave band. Differences are shown between the stimuli Speech and Speech SN6*

The difference in cut-off modulation frequency for speech and speech and noise is shown in figure 3-9. Note that the difference between absolute values of CR_{eff} between both stimuli as found in figures 3-4b and 3-4d is eliminated by taking the 75% value relative to the maximum CR_{eff} value. For the shortest release times the compressor is relatively more effective for the speech stimulus, while for longer release times the effectiveness is the same. The reason is most probably a disturbing effect of the noise floor in the modulation spectra at short modulation frequencies. Figures 3-4a and 3-4c show that the disturbing effect is stronger for speech-in-noise than for speech only.

Figure 3-10 shows the effect of the attack time on the effectiveness for both the Speech SN6 and the ICRA 1spSN6 stimuli. The release time has been kept constant to a value of 15 ms. A gradual decrease in the value of F_m cut-off is found for both stimuli. Just like with the release time we find higher cut-off modulation frequencies for the ICRA noise.

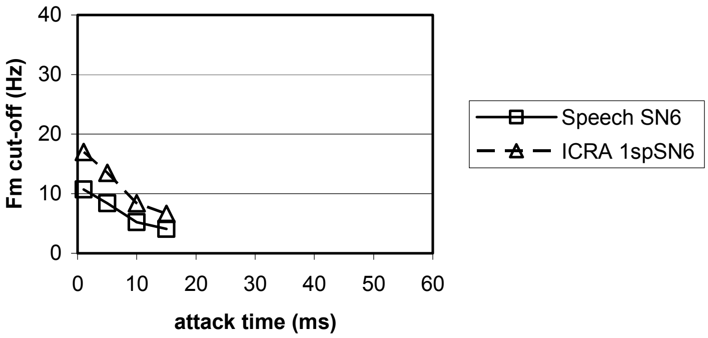


Figure 3-10 *75%- cut-off modulation frequency for CR_{eff} as function of attack time of the compressor, using CR=2 and a fixed release time of 15 ms within a 1-kHz octave band. Differences are shown between the stimuli Speech SN6 and ICRA 1spSN6.*

C. Frequency-dependent behaviour as function of channel properties

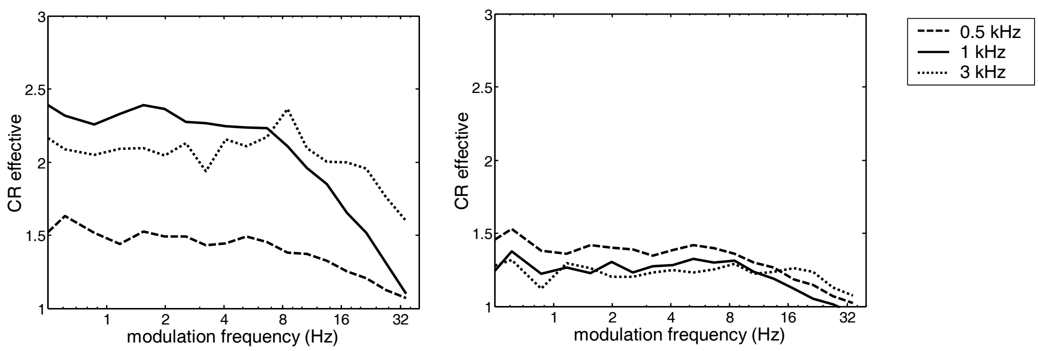


Figure 3-11a/b *Effective compression analysed within different octave bands, comparing 3-band compression with cross-over frequencies 703 Hz and 1359 Hz (11a) and single-channel broadband compression (11b) for the Speech SN6 stimulus. In all bands CR=2 is used with attack time of 1 ms and release time of 15 ms.*

The difference between three-channel compression and single-channel broadband compression are shown in figures 3-11a and 3-11b using analyses within octave-bands 500 Hz, 1kHz and 3 kHz respectively. As stimulus the Speech SN6 signal is used. For the multi-channel compressor a significant amount of compression is found within each of the three octave bands analysed. The compression is most effective within the 1-kHz and the 3-kHz band. The amount of effective compression even exceeds the theoretical value of 2, just like the results shown in section A. The broadband compressor mainly reduces the modulations within the low-frequency band as these modulations dominate the broad-band signal. Only a low amount of effective compression is found within the 1-kHz and 3-kHz band. The method nicely illustrates this fundamental difference between broadband and multi-channel compression.

Figures 3-12a and 3-12b show the results for the three-channel compression setting with CR=2 in each of the three bands for the ICRA 1spSN6 and the ICRA 6sp stimuli respectively. Although the most effective compression is found within the 1 kHz band, a significant amount of compression is also found within the low- and high frequency octave bands.

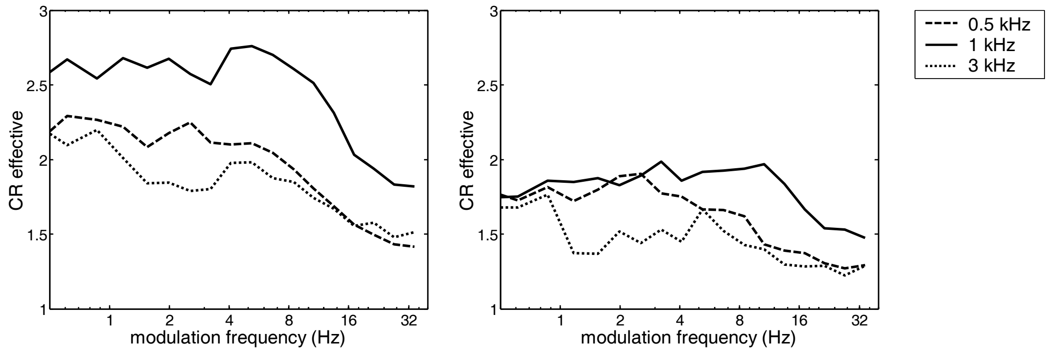


Figure 3-12a/b . *Effective compression analysed within different octave bands, using 3-band compression for the ICRA 1spSN6 stimulus (12a) and the ICRA 6sp stimulus (12b) with cross-over frequencies 703 Hz and 1359 Hz. In all bands CR=2 is used in combination with an attack time of 1 ms and a release time of 15 m.*

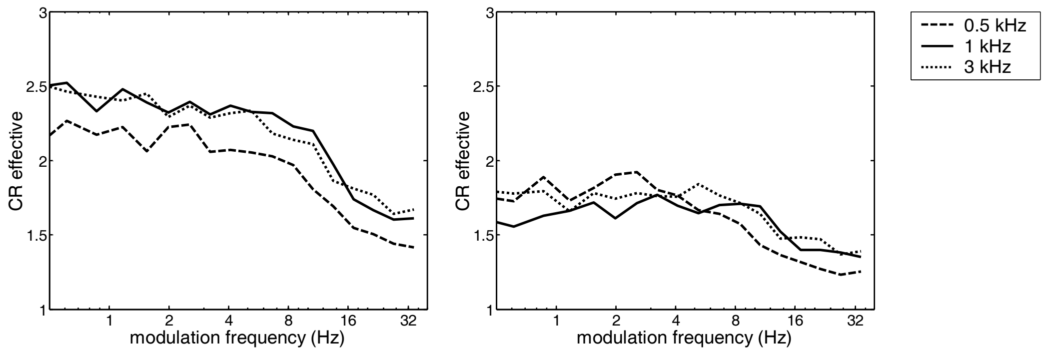


Figure 3-13 *Effective compression analysed within different octave bands, using 3-band compression for the ICRA 1spSN6 stimulus (13a) and the ICRA 6sp stimulus (13b) with cross-over frequencies 890 Hz and 2290 Hz. In all bands CR=2 is used in combination with an attack time of 1 ms and a release time of 15 ms.*

To investigate the effect of the cross-over frequencies, we also tested a three-channel configuration with similar cross-over frequencies as the modulation bands of the ICRA noise. Figures 3-13a and 3-13b show the effective compression ratio for the same stimuli and same compression settings as in figures 3-12a/b, except that the cross-over frequencies have been changed to 890 Hz and 2290 Hz. Now hardly any differences are found between the curves of each frequency band. This shows that the choice of cross-over frequencies influences the amount of effective compression within each frequency band. An important reason is that in the last example each compression channel is only driven by one independent stream of modulations, whereas in the previous example (figures 3-12a/b) there is probably interference between different streams of modulations within one compression channel. Another relevant factor is the broadness of the compression channels, as a smaller channel will generally result in a more effective compression (compare the results for the 1-kHz band in figures 3-12a/b and 3-13a/b).

3.4 Discussion

Methodological aspects

An analytic method has been tested that estimates the effective compression ratio in hearing aids for speech-(like) stimuli. The method compares the modulation spectra with and without processing, resulting in an effective compression ratio as function of modulation frequency. In general the method presents realistic and consistent results that depend of the stimulus used.

In section A we showed that the curves are characterised by an almost constant effective compression ratio at low modulation frequencies, followed by a drop in effectiveness at higher modulation frequencies. These kinds of curves are also found in measurements using amplitude-modulated signals (Verschuure et al. 1996, Stone et al. 1992). The amount of effective compression at low modulation frequencies roughly approaches the static compression ratio for the WDRC-setting. Above a certain frequency the modulations become too fast for the compressor, resulting in a less effective compression.

A remarkable difference in effective compression is found between speech and speech in stationary noise. With stationary noise we found values that exceeded the static compression ratio, while without the noise the effective compression remained below the static value. A reason might be the difference in modulation depth between both stimuli. For the highly modulating speech signal, we assume an increased influence of non-linear harmonics. This is caused by a definition of compression characteristics in dB instead of linear units. If the modulation depth exceeds a value of 0.45 significant harmonic products are introduced by the compressor in the case of an amplitude-modulated sine-wave (>5%, Verschuure et al. 1996). The formula used to estimate the effective compression ratio is not valid anymore. The resulting effect on the modulation spectrum is hard to predict. We expect that the effective compression ratio at modulation frequency F_m decreases because of a less effective reduction of the modulations at this frequency. The effective compression ratio at harmonic frequencies of F_m probably also decreases because the compressor introduces modulations at these frequencies that were not present at the input of the compressor. This would result in a less effective reduction of speech modulations and thus a lower CR_{eff} for a highly modulating stimulus as speech.

A similar difference is found between the ICRA stimuli with and without stationary noise. It is however remarkable that almost no effect is found of the number of simulated speakers. The curves found for the ICRA 6-speaker condition are very similar to those found for the ICRA 1-speaker condition. This seems in contradiction to our expectation that the amount of modulation is a critical factor. Different results are found for stimuli with a similar amount of modulation whereas similar results are found for stimuli with a different amount of modulation. Therefore the character and shape of the envelope function seems to be a more critical factor than the average amount of modulation. Visual inspection of the envelope functions reveals that the speech-in-noise stimuli are characterised by a regular shape of the envelope with only a few pronounced peaks while the multi-speaker conditions are characterised by a larger number of smaller peaks. This means that the speech-in-noise stimuli have a higher risk to locally exceed the critical value of m=0.45. Apparently this results in an increased value of the effective compression ratio for these stimuli. We have no logical explanation for his effect. The bias in estimated CR_{eff} at high modulation depth remains a topic of future research.

In section B we focussed on the temporal effects by defining a cut-off frequency above which the effective compression ratio starts to decline. We found a cut-off frequency of 10 to 20 Hz when using a release time of 15 ms and an attack time of 1ms. For higher release times the cut-off frequency quickly decreases to values of 2 to 4 Hz. We would expect to find slightly higher cut-off frequencies considering the small time windows used by the compressor. However, the structure of the speech modulations is rather complex compared to for instance amplitude-modulated sine-wave signals. Therefore the fastest speech modulations will be difficult to follow, even when the time constants should be fast enough in theory. It is encouraging that the temporal characteristic of the compressor as revealed by the method does hardly depend of the stimuli used. For this purpose the method seems to be rather robust. A small difference is that for the ICRA stimuli the compression is slightly more effective at high modulation frequencies compared to the real-speech stimuli. The difference is most probably caused by a difference in signal-to-noise ratio in the modulation spectrum at high modulation frequencies. This might also explain that we found a relatively more effective compression of speech compared to speech-in-noise when using very fast compression.

In section C we have investigated the effect of compression-channel properties on the effective compression within various frequency bands. The method succeeds to distinguish between single-channel broadband compression and multi-channel compression. In the latter case the modulations are effectively reduced within the low- middle and high-frequency band, whereas for broad-band compression we only found some effective compression within the low-frequency band. For multi-channel compression the choice of cross-over frequencies influences the amount of effective compression found within the various frequency bands. A first relevant factor is that in general the effective compression increases when using smaller compression channels. This explains that we found the most effective compression for the only 1-octave broad 1-kHz channel. A second reason is related to the specific character of the ICRA-stimuli, as they consist of three independent modulation-bands with fixed cross-over frequencies. When similar cross-over frequencies are chosen for the compression channels, almost no differences are found between the effective-compression values within the low-, middle- and high-frequency band. When the highest cross-over frequency is shifted downwards, the effective compression ratio within the high-frequency decreases considerably. The reason is that the high-frequency channel now contains two independent streams of modulations which cannot be both effectively reduced at the same time. This stimulus characteristic of the ICRA noises influences the results without being related to the effectiveness of the compression system.

It is clear that the choice of the stimulus does have an effect on effective compression ratio. This is not surprising as a certain interaction between stimulus and compressor settings is inevitable. As previously discussed, the shape of the envelope function and the distribution of the modulations over frequency bands determine the resulting effect of compression. This means that there is no such a thing as *the* effective compression characteristics of a certain system for speech(-like) signals in general. Important is that for each specific stimulus we found a realistic and consistent behaviour of the effective compression after adjusting certain compression parameters.

The question that remains is what specific stimulus should be chosen for further applications of the described method. The ICRA modulated noises have the advantage that they are well described and standardised. A disadvantage is the fixed cross-over frequencies of the frequency bands containing independent modulations. Another kind of choice should be made between 1-speaker, multi-speaker and speech-in-noise conditions. As discussed above, the multi-speaker condition has probably the lowest risk of exceeding the "critical" modulation depth of 0.45. Furthermore, we found realistic values of the effective compression ratio for these stimuli. Therefore a multi-talker condition like the 6-speaker ICRA noise seems a safe choice until we know more about the behaviour of CR_{eff} at high modulation depth. Another possibility that was not tested is to use a multi-speaker condition using real speech. Such a stimulus would combine the advantages of a relatively smooth modulation pattern and no fixed modulation bands.

Clinical implications

One of the strongest points of the present method is the detailed information that is obtained about the temporal behaviour of the compressor for speech-like stimuli. In nowadays hearing aids many different compression settings and systems are being used. Sometimes the hearing aid is assumed to reduce the speech modulations using so-called syllabic or phoneme compression systems. In other cases only the variations in overall-level of speech should be adjusted while the speech modulations should remain unaffected. We think it is relevant to know and verify whether the system behaves according to the expectations.

Our results show that a release time of less than 30 ms is needed to effectively compress most of the relevant speech modulations. When using a release time of for instance 50 ms, which is commonly used in syllabic compression systems, only a small part of the speech modulations are effectively compressed (<4 Hz). This is relevant information for clinicians that want to fine-tune the hearing-aid settings to the needs of a hearing-impaired individual.

With a standardised input signal the method described in this paper could be developed to a standard tool to measure the effect of compression on speech in modern complex hearing aids. The non-linear processing usually is a combination of many nonlinearities and the effect of compression on the speech signal is therefore hard to predict. A standardised method will be a valuable instrument to compare the resulting effect of non-linear processing in various hearing aids. However, an important prerequisite is that the noise suppression system in the hearing aid can be turned off as otherwise the effect of compression and noise suppression will interfere. A properly functioning noise suppression algorithm increases the amount of modulation, which counteracts the effect of compression in terms of modulation transfer.

3.5 Conclusions

A new analytic method has been introduced to estimate the effect of dynamic-range compression on speech or speech-like signals as function of modulation frequency. The results depend on the stimulus used. Both speech and artificially modulated noise (ICRA) seem to be appropriate stimuli. It has the potential to be used as a standardised method to characterise non-linear hearing-aid amplification for speech signals. The results show that the relevant intensity differences in speech are affected only by compression with relatively fast time constants (<30 ms).

3.6 References

American National Standards Institute (1996). Specifications of hearing aid characteristics. ANSI S3.22-1996 New York.

Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, Hagerman B, Heto R, Kei J, Lui C, Kiessling J, Kotby MN, Nasser NHA, El Kholy WAH, Nakanishi Y, Oyer H, Powell R, Stephens D, Meredith R, Sirimanna T, Tavartkiladze G, Frolenkov GI, Westermann S, Ludvigsen C (1994). An international comparison of long-term average speech spectra. J Acoust Soc Am. 96: 2108-2120.

Dyrlund O, Ludvigsen C, Olofsson A, Poulsen T (1994). Hearing aid measurements with speech and noise signals. Scand Audiol 23(3): 153-157.

Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. (2001) ICRA noises: artificial noise signals with speech-like spectral temporal properties fro hearing instrument assessment. International Collegium for Rehabilitative Audiology. Audiology 40(3): 148-157.

Goedegebure A, Hulshof M, Verschuure J, Dreschler WA. (2001) Effects of Single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners. Audiology 40:10-25.

Payton KL, Braida LD, Chen S, Rosengard P, Goldsworthy R. (2002) Computing the STI using speech as a probe stimulus. Published in Past, present and future of the Speech Transmission Index, Symposium TNO Human factors Soesterberg, editor v. Wijngaarden SJ (ISBN 90-76702-02-0): 125-138.

Priestley MB 1981 Spectral analysis and time series. Academic Press Harcourt Brace Jovanovich, Publishers ISBN 0-12-564922-3

Plomp R. (1983) "Perception of speech as a modulated signal" in 10th International Congress of Phonetic Sciences. Utrecht.

Rass, U., Steeger, G.H., (2000). “A high performance Pocket-size System for Evaluations in Acoustic Signal Processing,” Acta Acustica 86: 374-375.

Steeneken HJM, Houtgast T. (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67(1):318-326.

Stone MA, Moore BCJ. (1992) Syllabic compression: effective compression ratios for signals modulated at different rates. Brit J Audiol 26:351-361.

Versfeld NJ, Daalder L, Festen JM, Houtgast T (2000). Methods for the selection of sentence materials for efficient measurements of the speech reception thresholds. J Acoust Soc Am 107(3): 1671-1684.

Verschuure J, Maas AJJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA. (1996) Compression and its effect on the speech signal. Ear Hear. 17:162-175.

4 Effects of single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners.

Audiology (2001) 40(1):10-25
Goedegebure A, Hulshof M, Verschuure J, Dreschler WA

Abstract

The effect of digital processing on speech intelligibility was studied in hearing-impaired listeners with moderate to severe high-frequency losses. We varied the amount of smoothed phonemic compression in a high-frequency channel using wide-band control. Two alternative systems were tested to compensate for Upward-Spread-Of-Masking (USOM) and to reduce modulations in the high-frequency channel effectively. Consonant-Vowel-Consonant (CVC) tests were conducted in a group of 14 subjects, using 8 different speech-processing settings. Speech intelligibility improved significantly with compression, mainly due to positive effects on the initial-consonant score. Surprisingly, listeners with a smaller residual dynamic range tended to profit less from compression. Compensation for USOM gave an additional improvement of vowel intelligibility. In a background noise, we found consistently negative effects of speech processing. We concluded that the combined use of phoneme compression and USOM compensation is promising in conditions without background noise.

4.1 Introduction

It is almost impossible to imagine hearing aids today without dynamic-range compression. Most applications of compression are widely accepted, such as the use of slow compression as Automatic Gain Control (AGC) and the use of fast compression to suppress sudden uncomfortably loud sounds. More controversial is the use of fast compression to reduce the dynamics of speech, also indicated by "syllabic compression". A more appropriate term is phoneme compression, as the aim is to reduce the level differences between successive phonemes. Although many studies have investigated the expected benefits of phoneme compression, it is still controversial if phoneme compression should be used at comfortable levels well above threshold (Dillon 1996, Plomp 1988, Vilchur 199, Verschuure et al. 1993).

In the present study we focus on phoneme compression in one single high-frequency channel, using a similar system to that investigated by Verschuure et al 1993, 1996, 1998. With this system, improvements were found for consonant intelligibility in quiet and in a background noise with strong temporal fluctuations. Several other studies have also shown the advantages of single-channel compression compared to linear or multi-channel processing. Dreschler (1988) found slight positive effects on phoneme identification of wide-band compression with a low compression threshold. Bustamente and Braida (1987) showed that single-channel compression with high-frequency pre-emphasis in one channel is preferable to multi-channel compression. Moore and Glasberg (1988) found positive effects of fast-acting compression in a high-frequency channel on speech perception in background noise. Studies with multi-channel compression have hardly shown any benefits (Bentler and Nelson 1997, Bustamente and Braida 1987, de Gennaro et al 1986, van Harten-de Brijn et al 1997, Kollmeier et al 1993, Lippmann et al 1981, Marzinik et al 1997, Vilchur 1987, Walker et al 1984), except of Yund and Buckles (1995) who found significant improvements of speech perception at low Signal-to-Noise Ratio (SNR) with an 8-channel compression system.

The main rationale of single-channel phoneme compression is to improve consonant intelligibility by increasing the Consonant-Vowel Ratio (Hickson and Byrne 1997). If the compressor is controlled by the wide-band signal, the consonant cues will be emphasised as they contain generally much less energy than the vowels. This is the reason that Verschuure et al 1998 used the wide-band signal to control the compression within a high-frequency band. The improvements of consonant intelligibility found with the system indicate that the concept can be successful. However, some reasons against the use of wide-band control also can be mentioned:

1. The vowels contain most of the energy and will therefore receive reduced amplification by the compression system. Negative effects of Upward-Spread-Of-Masking (USOM) may become more prominent, especially if the compressor reduces the energy within the high-frequency band. Vowel intelligibility will not improve or can even decrease as a result.
2. The dynamics of speech in the high-frequency channel are not efficiently reduced by the compression system, because the low frequencies have more energy and therefore dominate the compression control. Level differences between consonants and vowels are reduced but the level distribution *within* the high-frequency channels stays about the same on average[6]. This means that a wide-band controlled compression system does not really compensate for the reduced dynamic range of the hearing-impaired listener.

The aim of the present study is to investigate two possibilities for improving the performance with the phoneme compressor of Verschuure et al 1998, by compensating for the two negative aspects of wide-band control described above. One suggested solution consists of extra low-frequency reduction at high levels to compensate for USOM. The second solution uses high-frequency emphasis of the compression control to reduce the speech dynamics at high frequencies. The two alternatives and the basic algorithm are evaluated by obtaining speech scores in a group of hearing-impaired listeners with moderate to severe sloping hearing losses. The performance of the listeners with the different compression settings will be related to their residual dynamic range. This is particularly interesting as the dynamics of the processed speech signal in the high-frequency range differs between the algorithms.

Research questions

Our research questions are:

- Can the performance with single-channel phoneme compression be improved by additional compensation for USOM?
- Can the performance with single-channel phoneme compression be improved by a more effective reduction of speech dynamics within the high-frequency channel?
- Can the measured effects of compression be related to the residual dynamic range and the hearing loss of the listeners?

4.2 Methods

Description of the system

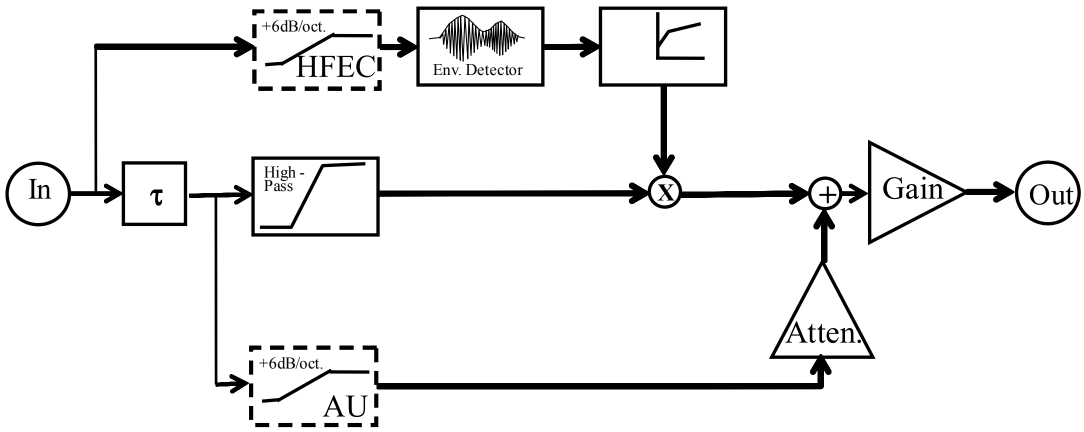
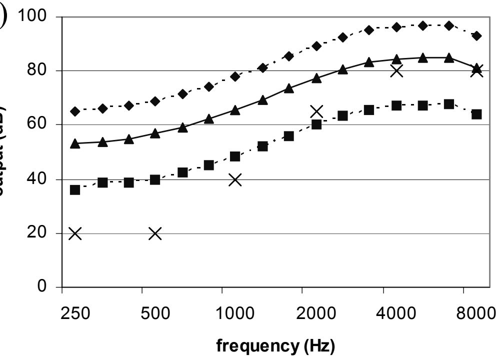


Figure 4-1 a/b/c *Simplified block scheme of the single-channel phoneme compressor. The dashed boxes represent the two +6 dB/octave filters tested in the conditions "Anti-Upward-spread-of-masking" (=AU, the filter in the linear path) and "High-Frequency Emphasis Control" = (HFEC, the filter in the control path).*

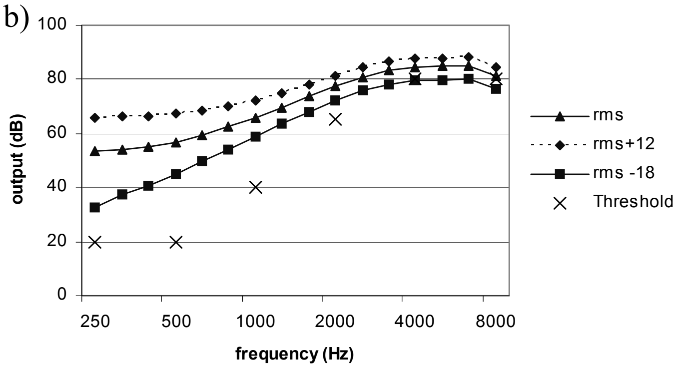
Figure 4-1 shows a simplified block scheme of the phonemic compressor. The basic configuration (without the dashed boxes) consists of three channels: a compression channel (in the middle) in which the high-frequency part of the input signal is processed, a control channel (on top) which calculates the required gain given the detected amplitude of the signal and a linear channel (below) in which the wide-band speech signal is attenuated and added to the compressed channel. An amplifier with adjustable gain determines the output-level. A delay τ is implemented between compression processing and compression control to reduce temporal distortion due to overshoots (Verschuure et al 1993).

The compression results in a level-dependant tilting of the frequency response as illustrated in Figures 4-2a and 4-2b. These figures show the third-octave band frequency responses of the system with the compression ratio set to 1 (figure 4-2a) and to 4 (figure 4-2b). The responses were obtained for pink noise presented at three different input levels: at the root-mean-square (rms) input level of speech and at levels 18 dB above and 12 dB below this rms level. Without compression the three frequency responses have the same ‘half-gain’ shape at distances +12 and -18 dB. This shape is adjusted to the pure-tone audiogram of each listener by choosing an appropriate combination of high-pass filtering in the compression channel and attenuation in the linear channel (see Appendix 4.9). When compression is switched on, the low-level frequency response gets steeper as the compression channel dominates the output signal. At a high level the response gets almost flat as the linear channel takes over the processing. For average speech levels the frequency response equals the half-gain shape as in the linear condition.

a)



b)



c)

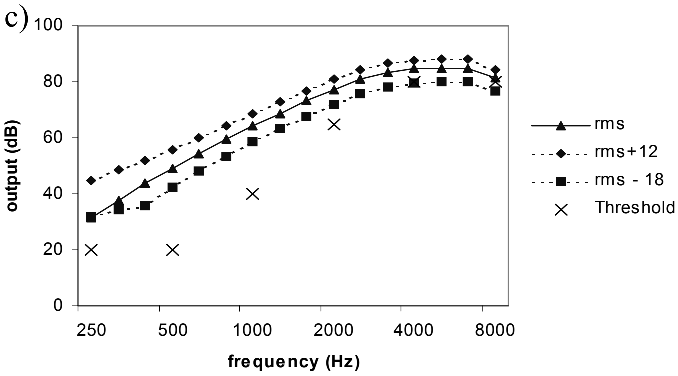


Figure 4-2 *Third-octaveband frequency responses of the system without compression (a), with compression ratio of 4 (b) and with compression ratio 4 and anti-USOM filtering (c). As input signal pink noise was used at three levels: rms of speech (triangles), rms+12 dB (diamonds) and rms-18 dB (squares). The system was tested with a typical setting of the mirrored audiogram-filter and attenuator values, corresponding to the audiogram depicted in the figure (crosses).*

The aim of the varying frequency response is to improve consonant intelligibility by increasing the audibility and reducing temporal masking of weak speech components at high frequencies. However, the flat frequency response at high levels induces more Upward-Spread-Of-Masking (USOM), which may have an adverse effect on speech perception. Compensation for USOM can be achieved by increasing the balance between high- and low-frequency amplification. Previous experiments showed that the use of emphasis in front or behind the system has negative effects on consonant intelligibility which counteract the positive effects on vowel intelligibility (Verschuure et al 1994). In the present study we want to avoid the reported negative effects on consonant intelligibility while maintaining the positive effects on the vowels. Therefore we implemented extra high-frequency emphasis in *the linear channel* indicated as 'anti-USOM filtering' (see the dashed box “AU” in figure 4-1). The filter has a positive slope between 500 Hz and 2.5 kHz of +6 dB\octave. Figure 4-2c illustrates the effect of the filter on the level-dependant frequency responses. For low input levels, the compression channel dominates the system output due to the high gain-values calculated by the compressor. The frequency response resembles that of the compression system without anti-USOM (figure

4-2b). For higher input levels the influence of the anti-USOM filter increases, which results in extra sloping of the response curves at low and middle frequencies.
A second alternative algorithm was realised by implementing a similar +6 dB/octave filter in *the control channel* (see the dashed box “HFEC” in figure 4-1). The filter almost flattens the speech spectrum, thus reducing the dominating effect of the low-frequency components in compression control. The speech dynamics in the high-frequency band will be more effectively reduced as a consequence. An additional effect is that the Consonant-Vowel-Ratio changes less than with wide-band control, because the filter reduces the level differences between consonants and vowels. The expected benefits on the consonant score may disappear because the Consonant-Vowel-Ratio is not increased enough or may even be reduced. This is the main reason why we did not implement complete high-frequency control.
Three types of compression were obtained by switching on and off the filtering in the linear channel or in the control channel: 'Wide-Band Control' (WBC) with no filtering in the linear path nor in the control path, 'Anti-Upward-spread-of-masking' (AU) with +6 dB/octave filtering in the linear path and 'High-Frequency Emphasis Control' (HFEC) with +6 dB/octave filtering in the control path. We evaluated each of the three compression algorithms in our speech-intelligibility experiments with compression ratios of 1, 2 and 4 respectively This resulted in eight different conditions: WBC1 (=HFEC1), WBC2, WBC4, AU1, AU2, AU4, HFEC2 and HFEC4.

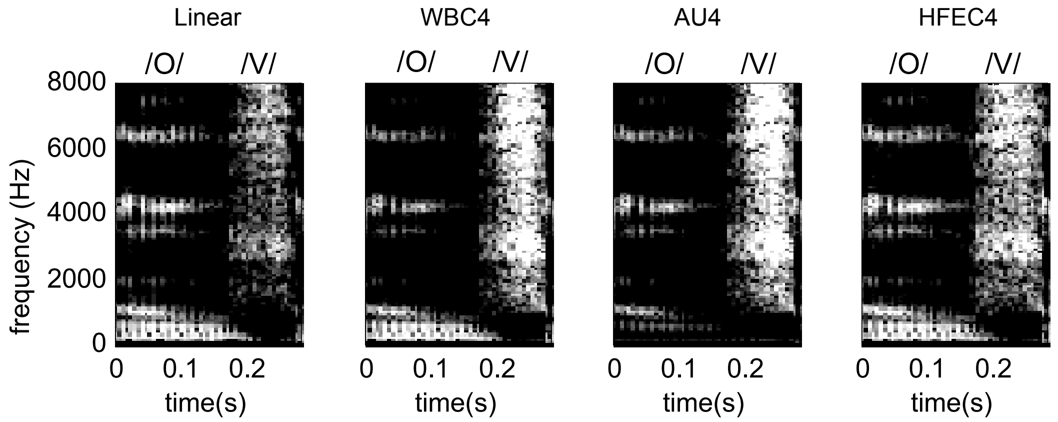


Figure 4-3a/b/c/d *Spectograms of the phonemes /O/ and /V/ in the dutch word 'over' with 4 different settings of the compression system: (a) is the system without compression, (b) is compression with CR=4, (c) is compression with CR=4 and anti-USOM and (d) is compression with CR=4 and high-frequency emphasis of the control signal.*

The short-term effects of the three types of compression on speech are illustrated by the spectrograms of figures 4-3a to 4-3d, showing the /o/ and the /v/ in the Dutch word ‘over’ (pronounced very similar to the English word) recorded with the system. Wide-band control (figure 4-3b) clearly increases the contrast between both phonemes by emphasising the /v/ and attenuating the /o/ compared to linear processing (figure 4-3a). With high-frequency emphasis control (figure 4-3d) the high-frequency energy is more equally distributed over both phonemes, indicating that the modulations at high frequencies are reduced by this type of compression. An additional advantage is that the higher formants of vowels are not suppressed, which is illustrated by the weak third fromant of the /O/ around 2 kHz. This formant is attenuated when using WBC compression (figure 4-3b), which may cause

problems with the recognition of the /O/. With HFEC-compression the third formant is *not* attenuated or even slightly emphasised (figure 4-3d). The effect of anti-USOM filtering is mainly visible in the low-frequency region of the /O/ (figure 4-3c). The second formant (near 1000 Hz) can be better distinguished from the first formant (near 500 Hz) due to the additional low-frequency suppression.
The long-term effect of the three types of compression on speech is illustrated in figure 4-4. It shows the speech spectrum at the output of the system measured over a 30-seconds period with different settings of the system. We used the same speech material and input level as used for the speech intelligibility measurements. The compressor settings were the same as used in figures 4-2 and 4-3. As expected, the differences are generally small except for the extra suppression of low frequencies by the anti-USOM algorithm. Above 1 kHz, a slight additional tilting of the frequency response is found for the two conditions with wide-band controlled compression (WBC and AU). This effect can be explained by the relative emphasis of consonant energy (mainly above 2 kHz) compared to vowel energy (mainly below 2 kHz).

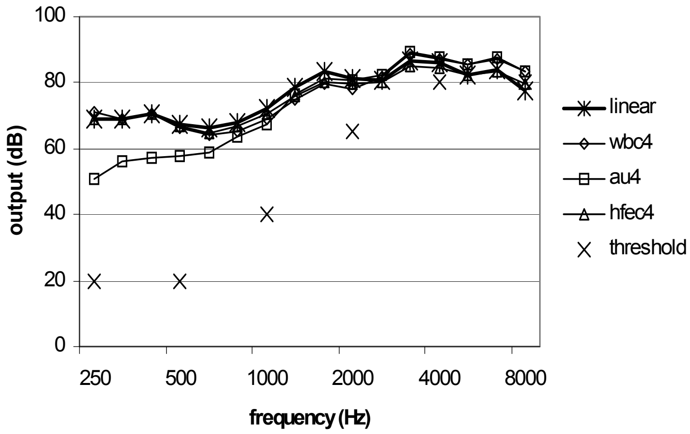


Figure 4-4 *Third-octave band frequency responses of the system averaged over a 30-seconds long sample of speech. The system was set to no compression (stars), compression with CR=4 (diamonds), compression with CR=4 and anti-USOM (squares) and compression with high-frequency emphasis of the compression control (triangles).*

Speech tests

Consonant-vowel-consonant (CVC) word lists were used as speech material in both experiments (Steeneken et al 1990) . Each list consists of 51 CVC words (logatoms) embedded in five different carrier phrases. All consonant-vowel combinations and all vowel-consonant combinations exist in the Dutch language. The test words were random combinations of initial consonants, vowels and final consonants. This resulted in both nonsense and meaningful words. Successive test words with two or more phonemes in common were rejected. In total 24 different word lists were used spoken by the same male speaker. The multi-talker background noise was a 4 minute-long 'restaurant noise' from a commercially available library on CD-disks (Sound Ideas series 3000). It had been chosen because of its realistic character and its speech-like spectrum. This track was also used by Verschuure et al 1998. The carrier phrase was presented on a screen with an open space for the CVC-word. The listener was instructed to repeat the word literally, even if it did not make sense, or if he/she only heard only part of the word. The word that was repeated by the listener was typed in by the experimenter and displayed to the listener on a monitor. The listener was asked to acknowledge the answer. After that the response was automatically stored.

Loudness scaling

We quantitatively assessed loudness perception in listeners using a kind of "loudness growth in ½-octave bands" (LGOB) procedure (Allen et al 1990). The loudness was measured with headphones in the unaided condition. As stimuli we used third-octave bands of noise (centre-frequencies 250, 500, 1000, 2000 and 4000 Hz) and one test sentence of the CVC-lists. First, the Highest Comfortable Level (HCL) was determined for each of the stimuli, defined as the maximum level at which the listeners were happy to listen to for a long time without the sound becoming disturbingly loud. After this, the dynamic range was estimated by taking the difference between the HCL and the pure-tone threshold. The minimum presentation level in the loudness-scaling procedure was set at one third of the estimated dynamic range above the threshold in order to reduce the number of stimulus presentations at low levels. The maximum presentation level was set at HCL+5 dB. The step size was adjusted to 10, 5 or 2 dB in order to obtain a minimum of 5 and a maximum of 10 different presentation levels for each stimulus. The order of the stimuli levels was randomised. The response of the listener consisted of one of the 7 pre-defined Categorical Loudness Units (CLU): inaudible=0, very soft=1, soft=2, comfortable=3, loud=4, very loud=5 and uncomfortably loud=6. In addition, listeners were allowed to render judgements between two adjacent categories. The measurement was repeated at least two times at each level, or three times at levels that gave a difference in response of 1 CLU or more. All ratings were averaged for each presentation level. The highest level with a rating less or equal to 3 ('Comfortable') was defined as the Most Comfortable Level (MCL) in the case of the speech stimulus. The responses were fitted by the logistic function:

$$S = \frac{1}{\frac{1}{6} + x_0 x_1^{-L}} \qquad (4.1)$$

with S the loudness as function of presentation level L (dB HL) and the fitting parameters x_0 and x_1. The function has a limited range between the two asymptotes 0 (inaudible) and 6 (uncomfortably loud). The slope of the curve at MCL (S=3) was calculated from the fitted parameters. The difference between the fitted slope and a linear approximation was small in most cases as a result of the limited number of presentations at low levels.

Equipment

The speech signal and the background-noise signal were played from Digital Audio Tape (DAT) with a PC controlling the presentation. We used a programmable attenuator to adjust the level of the background noise and thus obtained the desired signal-to-noise-ratio. The signals were added on an analogue board, and after performing anti-aliasing filtering (9 kHz cut-off) the signal was sent to one of the input channels of the Ariel DSP56001-board on which the compressor had been implemented. After AD-conversion the samples were processed by the digital sound processor (sample frequency 22.050 kHz) and again converted into an analogue signal. An audiometer (Madsen OB822) was used to adjust the output level.

The loudness scaling was performed with the same settings. The stimuli (speech and narrow-band noise) were played from DAT and the presentation levels could be adjusted by means of the audiometer. Both speech and loudness-scaling measurements were automated by a computer program, except that the presentation levels on the audiometer had to be performed manually by the experimenter.

Hearing impaired listeners

The listeners were selected from the patient database of the Audiological Centre. All recently measured clinical tone- and speech-audiograms were screened for suitable candidates. The selection criteria were:

- cochlear hearing loss (air-bone gap < 10 dB),
- maximum speech score below 95% in the routine speech audiogram,
- sloping pure-tone audiogram (difference in threshold of at least 20 dB between the averages at 250 and 500 Hz, and at 4 and 8 kHz),
- the not-measured ear should be at the most 30 dB better than the measured ear to avoid cross-talk,
- the age of the subjects should be under 70 years.

A number of 31 patients were selected to be divided into two almost equal groups with pure-tone averages at high frequencies (average over 2, 4 and 8 kHz) higher than 70 dB HL (group I) and lower than 70 dB HL. The patients were asked to come to the Audiology department of the Erasmus Medical Centre Rotterdam at a number of different days to perform a number of speech intelligibility tests of about 1 to 2 hours duration each. They were paid only for their travel expenses. 18 of the selected patients volunteered to participate in the experiments. One of the 18 selected patients was not able to perform a CVC-test and another patient was rejected because of a diagnosed conductive hearing loss at the time of the testing. Two of the sixteen remaining patients did not finish the complete test procedure because they found the experiments too exhausting and time-consuming. The full set of data was collected for 14 listeners, of which the audiograms are shown in figure 4-5. These data were used in the analysis. The definitive group of listeners were representative of the originally selected patient population with regard to the inclusion criteria.

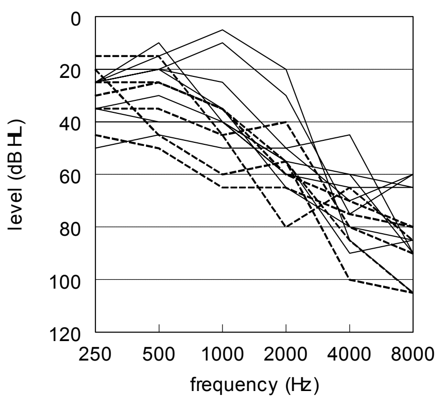


Figure 4-5 *Pure-tone audiograms of the tested ear of the listeners that participated in the listening experiments. The audiograms with the dotted lines correspond to listeners with a pure-tone average at high frequencies above 70 dB HL, the other audiograms to an average below 70 dB HL.*

Experimental design

The measurements were performed in four test sessions with one-week intervals for each listener. In the first session we first measured the pure-tone audiogram and the Most

Comfortable Level (MCL) for speech in the linear condition (see the loudness-scaling procedure described above). Then, two word lists were presented to the listener with four different settings of the compressor to acquaint the listeners with the task of repeating the CVC-words and with listening to compressed speech. Next, the output gain was optimised for maximum performance under linear processing by measuring the CVC-score at different output levels. First, the score was measured at MCL and at MCL+ 5 dB. If a better score was obtained at the higher level we also measured the score at MCL+10 dB, otherwise another measurement was performed at MCL-5 dB. The output gain that gave the highest score was used for all further CVC tests of a subject. In session 2, speech intelligibility in quiet was measured for each of the 8 conditions and loudness scaling was performed. The order of the test conditions was randomised using 8x8 Latin-square schemes while the word lists were presented in a fixed order. In this way, the combination of compressor condition with word list and with presentation order changed as much as possible for the different listeners. One word list was presented at the beginning of the session to reacquaint the listener with the task. The CVC-measurements of session 2 were repeated in session 3 with a new randomisation scheme and a different set of word lists. In session 4 we measured speech intelligibility in background noise. First, the Signal-to-Noise-Ratio corresponding to a CVC-score of 50% ($SNR_{50\%}$) was determined in the linear condition by varying the noise-level according to an adaptive staircase method as used by Verschuure et al.[18]. Then, the speech intelligibility scores were determined for each of the 8 compressor conditions at this $SNR_{50\%}$ level. A new randomisation scheme and a different set of word lists were used.
We used the statistical programme SPSS (version 7.5) for all our statistical analyses. Before using General Linear Model statistics we always verified that the difference scores between conditions across our listeners' population did not significantly deviate from a normal distribution (Kolmogorov-Smirnov 1-sample, $p>0.05$). The significance level of the statistical tests is defined by a p-value<0.05 if no other value is mentioned.

4.3 Results

Effects on CVC-score in quiet

First, we analysed the CVC-scores in quiet for learning effects by applying a two-way repeated-measures Analysis of Variance (ANOVA) with factors "time" (test- re-test) and "processing condition" (different settings of the compressor). No significant effects were found for "time" or for the interaction of "time" with "processing condition", indicating that learning effects were absent. We therefore used the results averaged over test and re-test sessions in the further analysis.
To analyse the effects of processing, we performed a two-way repeated measures ANOVA analysis with the factors "Compression Ratio" (CR=1, 2 or 4) and "Processing Type" (WBC, AU, or HFEC). The speech scores for conditions WBC1 and HFEC1 were identical in the analysis. The analysis was performed for the combined CVC-score, as well as for the initial-consonant, final-consonant and vowel scores separately. The significant effects found with the analysis are summarised in table 4-1, while figures 4-6a/d illustrate the size of the effects for the

individual processing conditions. Table 4-1 shows the significant main effects with corresponding F-values and significance level (*=p<0.05, **=p<0.01). Next, the significant contrast variables are given. The sign of the effect compared to the reference is indicated in the cells. If no significant main effect was found we did not test the corresponding contrasts. Figures 4-6a to 4-6d show the CVC-score, the initial-consonant score, the vowel and the final-consonant score respectively, averaged per condition as function of CR. The different types of processing are indicated by a different marker and line style.

Table 4-1 *Summary of repeated-measures ANOVA analysis with factors Compression Ratio "CR" and "Type of processing" on the phoneme score in quiet (* p<0.05, ** p<0.01, n.s. = not significant, n.t. = not tested).*

	CR	*Contrasts*			***Processing***	*Contrasts*		
		2-1	4-1	4-2		AU – WBC	HFEC -WBC	HFEC –AU
CVC	F(2,12)=5.19*	+ *	n.s.	n.s.	n.s.	n.t.	n.t.	n.t.
C1	n.s.	n.t.	n.t.	n.t.	n.s.	n.t.	n.t.	n.t.
V	n.s.	n.t.	n.t.	n.t.	F(2,12)=11.29**	+ *	+ **	n.s.
C2	n.s.	n.t.	n.t.	n.t.	n.s.	n.t.	n.t.	n.t.

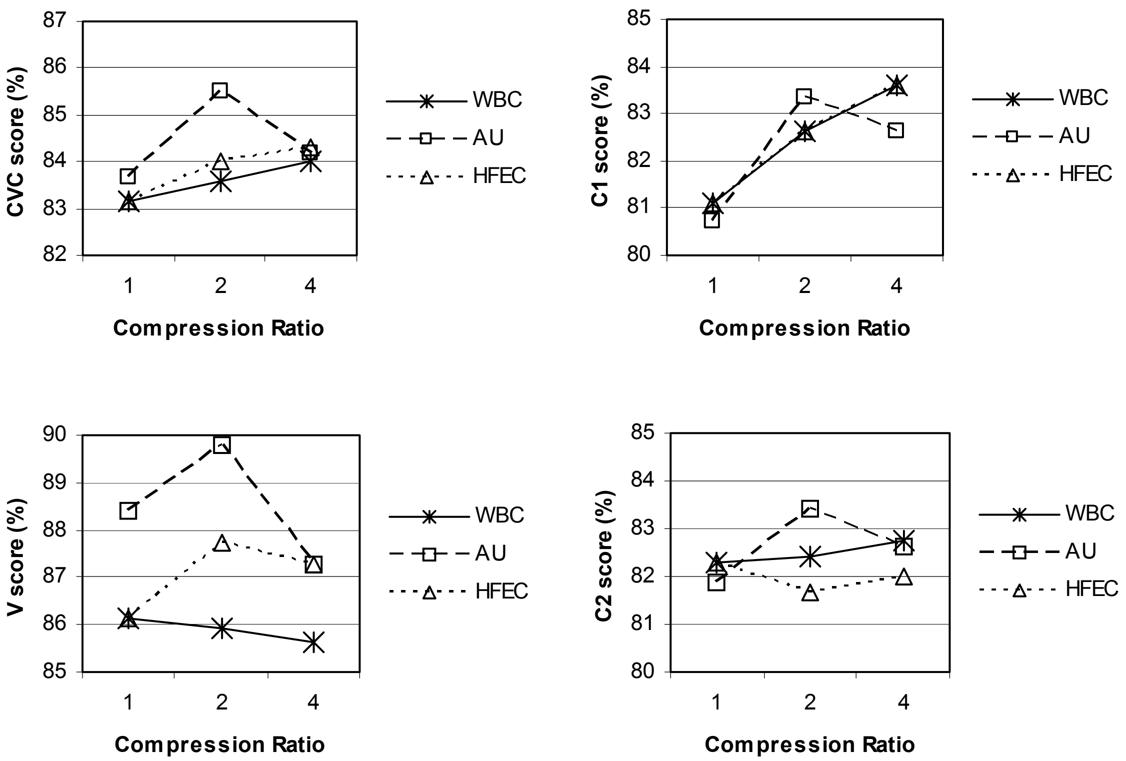


Figure 4-6a/b/c/d *The phoneme score as a function of Compression Ratio CR for each of the three processing conditions in quiet. The stars and the solid line represent the conditions with WBC, the squares and the striped line the conditions with AU , the triangles and the dotted line the conditions with HFEC. The results are shown for all items (a), only the initial consonant (b), only the vowel (c) and only the final consonant (d) respectively.*

A significant main effect was found for the factor Compression Ratio that can be attributed to a significant improvement with CR=2. This means that phoneme compression with a factor 2 has a significant positive effect on speech intelligibility in quiet, independent of the kind of implemented compression. Figure 4-6b shows that the difference is mainly explained by an improved initial-consonant score. Significant differences between the three different types of processing are found for the vowel score. Table 4-1 shows that *both* AU- and HFEC-compression significantly improve the vowel score compared to WBC compression. Figure 4-6c shows that vowel intelligibility stays about the same with WBC compression, whereas the score improves with HFEC and AU processing. It seems that both types of processing successfully compensate for spectral masking effects. Another interesting result is that no significant difference is found between HFEC and WBC for the consonant score. This means that consonant intelligibility does not improve by reducing the speech dynamics within the high-frequency channel more effectively. The final-consonant intelligibility gets even slightly poorer. Figures 4-6a to 4-6d show that the combination of AU-processing and Compression Ratio 2 gives the best performance. For this condition the improved initial-consonant score due to compression is combined with an improved vowel score due to anti-USOM. The improvement is significant positive compared to linear processing (paired T-test, $p<0.05$).

Effects on CVC-score in background noise

Table 4-2 *Summary of repeated-measures ANOVA of factors CR and Processing on the phoneme score in a background noise (* $p<0.05$, ** $p<0.01$, n.s. = not significant, n.t. = not tested).*

	CR	*Contrasts*			***Processing***	*Contrasts*		
		2-1	4-1	4-2		AU – WBC	HFEC -WBC	HFEC –AU
CVC	F(2,12)= 8.2 **	n.s.	-**	-**	F(2,12) = 4.2 *	- *	n.s.	+ *
C1	n.s.	n.t.	n.t.	n.t.	F(2,12) = 16.1**	-**	- *	- **
V	F(2,12)=11.8**	-**	-**	- *	F(2,12) = 5.8 *	n.s.	+ *	n.s.
C2	F(2,12)= 8.6 **	n.s.	- *	-**	F(2,12) = 4.5 *	- **	n.s.	- *

The effects found on speech intelligibility in background noise are almost opposite to results in quiet. Table 4-2 shows the results of a repeated-measures ANOVA analysis with CR and "Type of processing" as factors, while figures 4-7a/d illustrate the size of the effect for the different processing conditions.

Compression has a significant negative effect on speech intelligibility, mainly because of significant *poorer* results with CR=4. For CR=4 significant negative effects are found compared with both CR=1 and CR=2. The use of a compression ratio higher than 2 seems to be unfavourable in critical signal-to-noise ratios. Even more remarkable is the strong negative effect of anti-USOM processing on speech intelligibility in background noise. The listeners score about 7% poorer with the anti-USOM processing compared to the other two types of processing (see figure 4-7a). We expected the anti-USOM processing to reduce the negative effects on vowel recognition also in a background noise as vowels represent the louder parts of the speech signal, but unfortunately the opposite is true! Substantial negative effects on both vowel and consonant scores are found for all anti-USOM conditions. The HFEC processing does not show this large negative effect. We find even a significantly improved vowel intelligibility compared to compression with wide-band control. This means that high-frequency control compensates for undesired USOM effects caused by compression, even in background noise. Unfortunately, the resulting effect compared to linear processing is still negative. The only positive effect of compression compared to linear processing is found for the initial-consonant score with condition WBC2 (see figure 4-7b). The effect is however counteracted by a large negative effect of WBC compression on the vowel score. Therefore, phoneme compression seems of not much use in multi-talker babble background noise at critical signal-to-noise ratios.

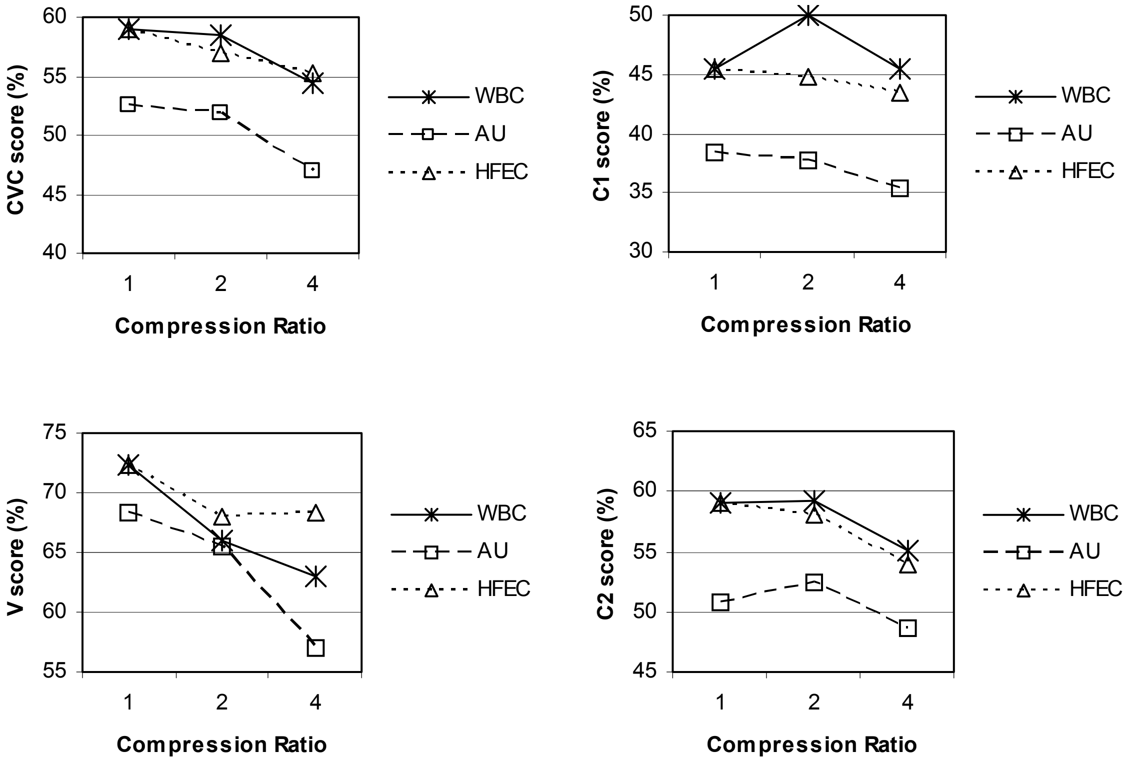


Figure 4-7a/b/c/d *The phoneme score as a function of Compression Ratio CR for each of the three processing conditions in background noise. The stars and the solid line represent the conditions with WBC, the squares and the striped line the conditions with AU , the triangles and the dotted line the conditions with HFEC. The results are shown for all items (a), only the initial consonant (b), only the vowel (c) and only the final consonant (d) respectively.*

It should be noted that the score is 59% with linear processing instead of the 50% target-score. This difference is most probably due to some acclimatisation to the processing and the task to repeat the CVC-words in background noise. Another reason might be that the shallow psychometric curve of non-sense CVC-words is not ideally suited for adaptive measurements. However, a 59%-level is still a critical listening condition in which effects of speech processing should be relatively large.

Relation to pure-tone audiogram and dynamic range

Next, we analysed the relationship of the observed effects on speech intelligibility to the differences in hearing loss and residual dynamic range *between* listeners. For this aim, we tested a number of between-subject variables on significant interactions with the two within-subject factors "Compression Ratio" and "Processing Type" in a repeated-measures MANOVA analysis. We considered the following variables: the average of the pure-tone audiogram at frequencies 2 and 4 kHz ($PTA_{2,4}$), the maximum difference between high-frequency (2 and 4 kHz) and low-frequency thresholds (250 and 500 Hz) of the audiogram (PT_{HLdif}), the slope of the audiogram (PT_{slope}), the dynamic range averaged over frequencies 2 and 4 kHz ($DR_{2,4}$) and the fitted slope of the loudness growth function averaged over frequencies 2 and 4 kHz ($Lslope_{2,4}$).

Table 4-3 *Significant interactions of between-subject variables and within-subject factor processing in a repeated-measures MANOVA for phoneme scores in quiet (* p<0.05, ** p<0.01, n.s.= not significant).*

phoneme condition	*between-subject variable*	*Processing*	*contrasts(to WBC)*	
			AU	*HFEC*
CVC	PT_{Hldif}	$F(2,11) = 20.26^{**}$	- **	n.s.
CVC	$DR_{2,4}$	$F(2,11) = 5.33^{*}$	+ *	n.s.
C1	PT_{Hldif}	$F(2,11) = 12.09^{**}$	- *	n.s.
V	PT_{Hldif}	$F(2,11) = 12.60^{**}$	- **	n.s.

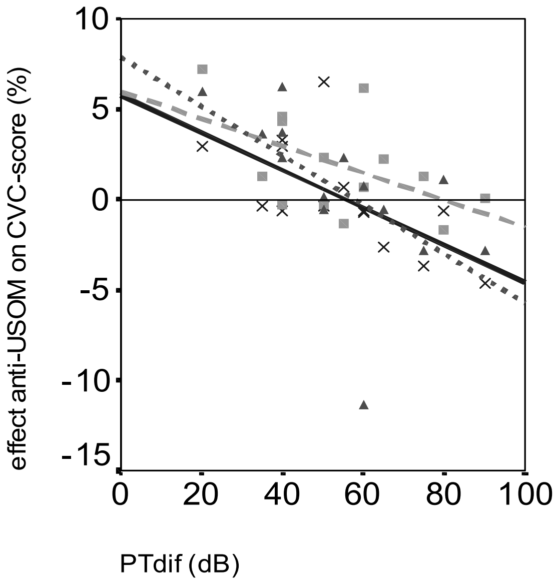


Figure 4-8 *The effect of anti-upward-spread-of-masking on the CVC-score in quiet as function of the difference between low- and high-frequency thresholds (PT_{dif}) for each of the listeners. The markers represent the difference in overall CVC-score for different compression ratios (1=linear, 2 and 4) between the conditions with anti-USOM (AU1=triangles, AU2 =squares and AU4 = crosses respectively) and the corresponding compression conditions without anti-USOM*

The significant interactions for the CVC-scores in quiet are listed in Table 4-3. The interaction with the contrast-variables and the sign of the correlation are shown in addition. The table shows that most significant interactions are found between the low-high-frequency contrast in the audiogram (PT_{HLdif}) and anti-USOM processing in quiet. The negative sign means that the listeners with relatively large difference between low- and high-frequency losses profit less from anti-USOM processing. This is illustrated by figure 4-8, which shows the effect of anti-USOM (i.e. the AU-conditions minus the corresponding WBC-conditions) for CR=1, 2 and 4 as function of the low-high-frequency contrast of the audiogram. Positive effects of anti-USOM processing are mainly found below values of 60 dB for PT_{HLdif} . The figure also shows the fitted linear regression lines which correspond to significant correlations of R=-0.59, R=-0.52 and R=-0.67 respectively.

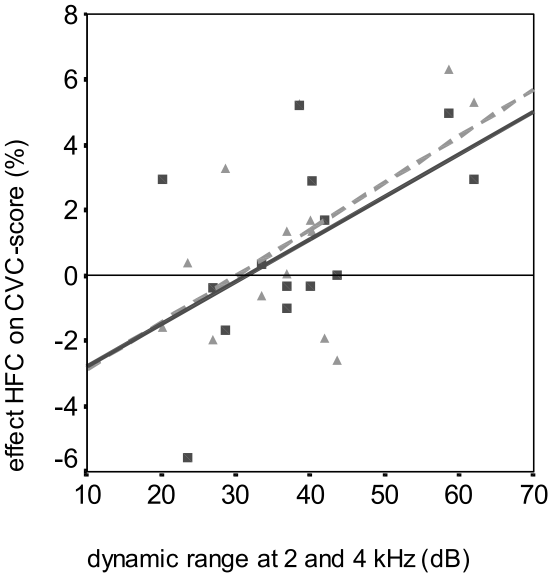


Figure 4-9 *The effects of high-frequency controlled compression on the CVC-score in quiet as function of the dynamic range at high frequencies (DR_H) for each of the listeners. The markers represent the difference in overall CVC-score for compression ratios 2 and 4 between the conditions with high-frequency emphasis control (HFEC2 = squares and HFEC4 = triangles respectively) and the linear-reference condition. The lines represent the fitted correlation functions (solid line and R=0.54 for HFEC2, dashed line and R=0.56 for HFEC4).*

The lack of significant interactions between "CR" and any of the factors is highly remarkable. It seems that the effects of compression are quite independent of the dynamic range, loudness growth or the pure-tone audiogram at high frequencies. This is contrary to expectations if we had assumed that compression would make more information available in a narrower dynamic range. Figure 4-9 illustrates that even the opposite seems true. It shows the effect of HFEC compression with CR=2 (triangles) and CR=4 (squares) on the CVC-score in quiet as function of the dynamic range at frequencies 2 and 4 kHz ($DR_{2,4}$). The correlations are not very strong but significant (R=0.54 and R=0.58 respectively, $p<0.05$). They show a trend toward mainly positive effects for listeners with dynamic ranges between 30 dB and 50 dB and toward negative effects for the listeners with a dynamic range below 25 dB. It seems that a dynamic range of 25 to 30 dB is the minimum required to obtain in quiet benefit from fast-acting compression.

In background noise, we only found a significant interaction between the fitted slope of the loudness function and the effects of CR on vowel intelligibility ($F(2,11)= 6.125$,$p<0.05$). This effect is due to a positive correlation between the slope of the loudness growth function and the effects of CR=2, which suggests again that the performance with compression gets better for the listeners with less recruitment. Also interesting is that no significant interactions are found with the effects of processing in background noise. The large negative effect of anti-USOM on speech intelligibility in noise does not depend of the hearing loss or the residual dynamic range of the listeners.

4.4 Discussion

The effects of compression on speech intelligibility in quiet meet the expectations fairly well. We found a significant positive effect of phonemic compression with CR=2 that can mainly be attributed to an improved initial-consonant intelligibility. This is according to our assumption that single-channel compression increases the audibility of weak consonant cues by increasing the Consonant-Vowel Ratio (CVR). Surprisingly, reduction of forward masking seems to play no important role as no improvements were found for the final-consonant score. It seems that for the final consonant increased audibility is counteracted by a second negative effect. Dreschler (1988) mentions temporal distortions of consonants as a possible negative side-effect of phoneme compression. In our system, however, we compensated for temporal distortion by reducing the overshoots. Hickson and Byrne (1997) suggest that the CVR it self is an important cue which may be affected by single-channel phoneme compression. If this is true, the final-consonant intelligibility should be improved by high-frequency emphasis of the compression control as this type of compression causes less changes in the CVR. We did not find such an improvement, so either CVR-cues do not play an important role or there is a trade-off effect between increasing audibility and affecting CVR-cues.

Next, we found that anti-USOM filtering and high-frequency emphasis of the compression control both have a significant positive effect on vowel intelligibility. One of the purposes to implement both types of compression was to reduce spectral masking effects. Interestingly, reducing the energy of low-frequency formants (AU-processing) or enhancing the weak high-frequency formants (HFEC-processing) have both the desired effect on vowel intelligibility. This means that vowel intelligibility is mainly determined by the balance between high- and low-frequency gain and less by the absolute gain levels. Apparently, the balance of the applied linear half-gain frequency shaping was not suited to completely compensate for USOM. A steeper frequency response seems required to optimise vowel intelligibility for listeners with moderate to severe high-frequency losses as found in our subject population.

Maximum benefit for patients was obtained with two separate channels for compression and anti-USOM filtering. This combination successfully enhances both temporal and spectral cues, apparently without the undesired interactions as found in an earlier study of the use of anti-USOM filtering implemented as pre- and post-processing (Verschuure at al 1994). The wide-band control compressor performs a double task in this condition: emphasis of weak consonants and activation of the anti-USOM filter in the case of vowels.

It should be realised that the small improvements found in quiet can still be very meaningful. First of all, the effects are measured in comparison to a very similar reference condition in which the presentation level is optimised with regard to speech intelligibility. This means that the measured benefit can really be attributed to effects of fast-acting speech processing and not to differences in frequency-shaping or presentation level. The effects are not due to improved speech intelligibility scores averaged over a wide range of levels, which is an effect one gets additionally. A second important aspect is the presence of a ceiling effect caused by a reasonable good performance in the linear condition. The overall phoneme score in quiet was 83% on average (± 9 %), whereas normal-hearing listeners usually score in between 95 and 100%. This means that measured improvements of 2 to 3 % on the score reduce the deficit in speech discrimination with 15 to 20%. This will constitute a noticeable improvement in performance for certain patients. A third point of interest is that positive effects are consistently found for all speech-processing conditions in this study and for similar conditions in previous studies. The effect of our system on speech intelligibility turns out to be very robust in quiet and can be logically explained from the measured effects on phoneme level.

Contrary to the improved performance in quiet, we found very poor results in background noise for the combination of anti-USOM and wide-band control. The negative effects can be partly attributed to the use of phoneme compression, as we found a significant negative overall effect for compression with compression ratio 4. According to Dillon (1996) it is not to be expected that phoneme compression improves speech intelligibility in noise. Compression within any frequency channel will affect the signal and noise within that channel by identical amounts. Moore et al (1995) and Verschuure et al (1998) suggested that the use of fast-acting compression might only show a benefit if the envelope of the background noise fluctuates markedly. The multi-talker babble noise in this experiment can be considered as a continuous background noise with a speech-like spectrum and only limited temporal fluctuations. Verschuure et al 1998 found positive effects on the initial-consonant score with the same background noise, resulting in a small positive overall effect. In our case, the negative effects on vowel intelligibility dominate the overall results. It seems that the reduction of vowel energy in the high-frequency channel has serious negative consequences for our group of listeners. The pure-tone audiograms of our listeners have somewhat steeper slopes in comparison to those of Verschuure et al 1998, which may be the reason for the poorer performance.
Most striking is the fact that the use of anti-USOM results in considerable negative effects on speech intelligibility in background noise, where we expected positive effects. To explain this we must realise that the present group of listeners has left less residual hearing capacities at high frequencies than at low frequencies. At a critical SNR, the transmission of high-frequency speech information will be seriously deteriorated while low- and middle frequency cues can still be recognised. It seems that anti-USOM filtering suppresses those speech cues that the listeners depend most on in noisy conditions. This could also explain our finding that listeners with steeper audiograms perform significantly worse with anti-USOM filtering (see table 4-3 and figure 4-8). On one hand we would expect benefits from USOM as such hearing losses are normally characterised by relatively broad auditory filter-shapes at high frequencies. On the other hand, low- and middle-frequency speech cues are of major importance for these listeners as the system has a very poor spectral and temporal resolution at high frequencies. Therefore, any reduction of the relevant low- and middle frequency cues will seriously affect speech intelligibility. This suggests that anti-USOM processing by reducing the low frequencies should be avoided for listeners with relative steep audiograms and in critical speech-in-noise situations.
High-frequency emphasis of the compression control showed to be a better alternative to compensate for spectral masking effects in background noise. A significant positive effect on vowel intelligibility was found compared to the use of wide-band control. The results in quiet showed that the high-frequency emphasis control can be considered as a kind of anti-USOM processing. It has the advantage that the lower formants are not affected, which is most probably the reason for the relatively good performance in background noise. Unfortunately, the initial-consonant perception significantly decreases as compared to wide-band control compression. Consequently, the overall performance with high-frequency control and wide-band control is about the same in background noise.

There is another interesting aspect to the results. The improvement of speech intelligibility is sometimes attributed to the ability of compression systems to match the dynamics of speech into the reduced dynamic range of the impaired ear. Our results, however, indicate that this is not the case here. Patients with a very small dynamic range benefited less from phonemic compression that reduces the speech dynamics at high frequencies. The best results were found for patients with a dynamic range at high frequencies between 30 and 50 dB (corresponding to thresholds of about 50 to 70 dB HL). The smaller the dynamic range is, the less positive the effect of our compression. The emphasis of certain weak consonant cues seems to be a more important factor than fitting of the dynamic speech range in the dynamic hearing range.
As a consequence, the amount of single-channel phoneme compression should not be adjusted to the dynamic range of the individual. We found that a compression ratio of 2 is generally more appropriate than a value of 4, especially if background-noise situations are also considered. A high amount of single-channel compression seems to interfere with the detection of speech cues in background noise, independent of the residual dynamic range of the listener. Only in case of many independent compression channels it may be beneficial to fit the compression ratios to the individual dynamic range (Yund and Buckles 1995). The same accounts for slow compression (time constants of at least 500 ms) as it does not affect the fast modulations of speech. Therefore, single-channel phoneme compression should always be fitted with low CR-values. Additional slow compression should be used to compensate for the reduced dynamic range of the individual hearing loss.
It can be argued that the limited benefit of compression for the listeners with the more severe high-frequency losses may have been influenced by insufficient audibility of the speech spectrum. In our experimental design we chose to optimise the reference speech spectrum for intelligibility instead of guaranteeing complete audibility. This means that the high-frequency tail of the spectrum may have dropped below threshold for the steeper hearing losses. It is however unlikely that this played an important role in the measured effects of compression. First of all, the optimisation for speech intelligibility implies that eventually non-audible parts did not significantly contribute to speech perception if they were made audible. Secondly, the high-frequency losses are generally not too severe (65 dB HL or lower at 2 kHz, except for one listener), so at least an important part of the compressed high-frequency speech signal was presented above threshold. In the third place, compression has not much effect on the average spectrum at high frequencies (see figure 4-4). Only the audibility of *weak* high-frequency components may be increased, but that is one of the expected benefits of compression that should be measured for all kind of high-frequency hearing losses.

4.5 Conclusions

- the use of smoothed phoneme compression with compression ratio 2 in a high-frequency channel has a significant positive effect on speech intelligibility in quiet. In conditions with a multi-talker babble background noise we found mainly a negative effect of compression on speech intelligibility. Significant negative effects were found with compression ratio 4.

- the use of anti-USOM filtering and high-frequency emphasis control both significantly improve vowel perception in quiet in comparison to compression with wide-band control. In a multi-talker background noise, high-frequency emphasis control also significantly improves vowel intelligibility whereas remarkable significant negative effects are found for anti-USOM processing.
- listeners with a small dynamic range at high frequencies (smaller than 25-30 dB) benefited less from phoneme compression that reduces the speech dynamics at high frequencies than the listeners with a wider dynamic range.
- listeners with more than 60 dB difference between high- and low-frequency thresholds in the pure-tone audiogram benefited less from anti-USOM filtering than those with less difference.

4.6 Acknowledgements

This research has been performed in the framework of the project “Hearing Aid Research using Digital Intelligient Processing” (HEARDIP), which was part of the Technological Innovative programme for the Disabled and Elderly (TIDE) and was funded by the European Commission. We want to thank the partners of the Academical Medical Centre Amsterdam (Dept Audiology), the University of Oldenburg (Dept Medical Physics), the University of Cambridge (Dept of Experimental Psychology), Philips-Eindhoven and Siemens-Erlangen for their comments and contributions to our work. We want to direct our special thanks to Helene van Harten-de Bruyn, Peter-Paul Boermans and Sidonne van Kreveld-Bos of the AMC Amsterdam for the intensive discussions and their contributions to the development of our system. Furthermore, we would like to thank Chris Talens and René de Jong for their efforts in DSP programming. We also acknowledge the contributions made by the Heinsius Houbolt Foundation to support this research. Last but not least, we want to thank all hearing-impaired listeners that participated in the experiments. This work could not have been performed without their enthusiasm and persistence.

4.7 References

Allen JB, Hall JL, Jeng PS (1990) Loudness growth in 1/2-octave bands (LGOB)-A procedure for the assessment of loudness. J Acoust Soc Am 88 (2):745-753.

Bentler RA, Nelson JA (1997) Assessing release-time options in a two-channel AGC hearing aid. J Am Acad Audiol 6:43-51.

Bustamante DK, Braida LD (1987) Principal-component amplitude compression for the hearing impaired. J Acoust Soc Am 82:1227-1242.

Dillon H (1996) Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response time? Ear Hear 17 (4):287-307.

Dreschler WA (1988) The effects of specific compression settings on phoneme identification in hearing-impaired subjects. Scand Audiol 17:35-43.

de Gennaro S, Braida LD, Durlach NI (1986) Multichannel syllabic compression for severely impaired listeners. J Rehabil Res Dev 23:17-24.

van Harten-de Bruijn HE, van Kreveld-Bos CSGM, Dreschler WA, Verschuure J (1997) Design of two syllabic non-linear multi-channel signal processors and the results of speech tests in noise. Ear Hear 18:26-33.

Hickson L, Byrne D (1997) Consonant perception in quiet: effect of increasing the Consonant-Vowel Ratio with compression amplification. J Am Acad Audiol 8:322-332.

Moore BCJ, Glasberg BR (1988) A comparison of four methods of implementing automatic gain control (ACG) in hearing aids. Brit J Audiol 22:93-104.

Kollmeier B, Peissig J, Hohmann V (1993) Real-time multiband dynamic compression and noise reduction for binaural hearing aids. J Rehabil Res Dev 30:82-94.

Lippmann RP, Braida LD, Durlach NI (1981) Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. J Acoust Soc Am 69:524-534.

Marzinzik M, Hohmann V, Appel JE, Kollmeier B. (1997) Evaluation of different multi-channel dynamic compression algorithms with regard to recruitment compensation, quality and speech intelligibility. in Seventh Oldenburg symposium on psychological acoustics. In: Schick A, Klaate M, eds. Contributions to psychological acoustics- results of the 7th Oldenburg symposium on psychological acoustics, Oldenburg: 619-630

Moore BCJ, Glasberg BR, Vickers DA (1995) Simulation of the effects of loudness recruitment on the intelligibility of speech in noise. Brit J Audiol 29 (3): 131-143

Plomp R (1988) The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function. J Acoust Soc Am 83:2322-2327.

Steeneken HJM, Geurtsen FWM, Agterhuis E (1990) Speech data-base for intelligibility and speech quality measurements, TNO-Institute for Perception: Soesterberg.

Verschuure J, Dreschler WA (1993) Present and future technology in hearing aids. JSPLA Monogr; (Suppl.):65-73.

Verschuure J, Dreschler WA, de Haan EH (1993) Syllabic compression and speech intelligibility in hearing-impaired listeners. Scand Audiol S38: 92-100.

Verschuure J, Prinsen TT, Dreschler WA. (1994) The effects of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people. Ear Hear 15:13-21.

Verschuure J, Maas AJJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA (1996) Compression and its effect on the speech signal. Ear Hear 17:162-175.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Villchur E. Multiband compression processing for profound deafness. J Rehabil Res Dev 1987; 24:135-148.

Villchur E. (1989) Comments on "The negative effect of amlitude compression in multichannel hearing aids in the light of the modulation-transfer function". J Acoust Soc Am 86(1):425-427.

Walker G, Byrne D, Dillon H (1984) The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. J Acoust Soc Am 76(3):746-757.

Yund EW, Buckles KM (1995) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239.

4.8 Appendix: Implementation of the system

A detailed block scheme of the system is shown in figure 4-10. The envelope detector is composed of a rectifier followed by a switching low-pass filter. The filter is a first-order filter with an attack time of 5 ms and a release time of 15 ms, based on results in previous experiments (Verschuure et al 1994). The output level of the envelope detector is used as input level for a look-up table which calculates the corresponding gain by means of an interpolation algorithm. We used three different compression tables in the experiments defined by compression ratios 1, 2 and 4. A compression range of 45 dB was used for each of the tables, which is enough to perform a full-dynamic-range compression of the speech material for the tests. Linear processing was used below the compression threshold. The three tables intersect at an input level of –22 dB versus maximum signal. The root-mean-square input level of the speech was adjusted to the –22 dB level, in order to keep the average output level and the average frequency response of speech constant for the three different compression ratios. The delay τ is implemented to reduce temporal distortion due to overshoots. The delay is set to 3 ms, based on the results in speech intelligibility tests (Verschuure et al 1993) and experiments on the effectiveness of the compressor in modulation reduction (Verschuure et al 1996) . The first cross-connection between the linear path and the compression path is used to reduce the stop-band ripple of the Mirrored Audiogram filter (MA). The attenuation factor Att1 is adjusted to the difference in hearing loss between low and high frequencies for each listener. This corresponds to the setting with the maximum difference in gain between the low and the high frequencies. The frequency response is set to half-gain for rms-input level by adjusting the attenuation factor in the second cross-connection (Att2) to half the value of Att1. The match to the half-gain response is optimised by fine-tuning of the filter and attenuation settings based on visual inspection of the measured frequency response.

The implementation of both new filters is also shown in figure 4-10. The filtering in the control path causes a loss of low-frequency energy which is compensated with a gain factor G_{cc} of 12 dB. This means that on average the frequency response with filtering in the control path is similar to that without filtering. No level compensation is used in the linear channel because the anti-USOM filter reduces the low frequencies only in the case of relative loud speech parts. Relevant low-frequency speech information that is suppressed by the anti-USOM filter (mainly first-formant information) will still be audible to the listeners. All filters were implemented as Finite Impulse Response (FIR) -filters to avoid non-linearity in the phase characteristics of the signals in both the compression and linear channel. Extra time delays were implemented parallel to the FIR-filters to compensate for the time delay due to filtering (τ_{MA}, τ_{cc}, τ_{USOM}).

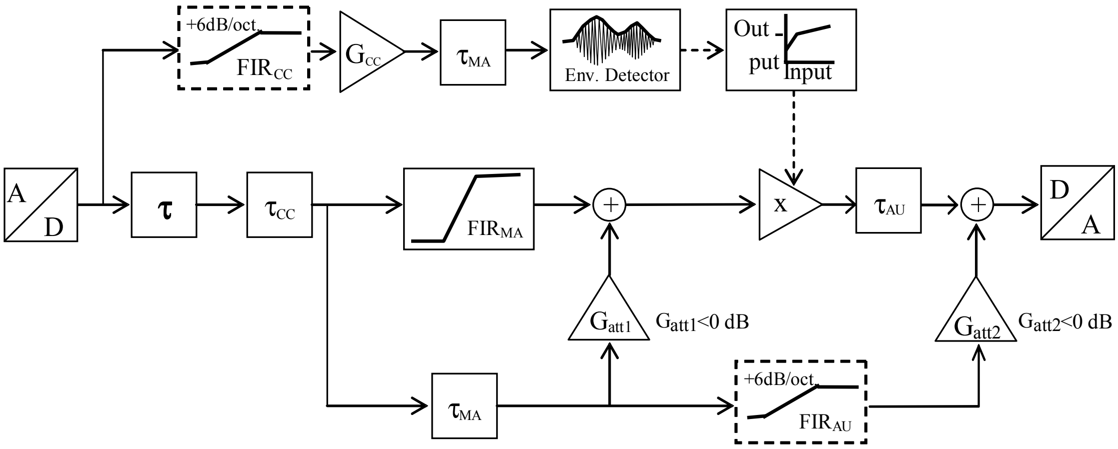


Figure 4-10 *Block diagram of the complete speech-processing system. The dashed boxes represent the +6 dB/octave filters tested in conditions AU and HFEC.*

5 The effects of phonemic compression and anti-upward-spread-of-masking (anti-USOM) on the perception of articulatory features in hearing-impaired listeners

Int. J. Audiol. (2002) 41(7): 414-428
Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA

Abstract

The effect of speech processing on articulatory-feature perception is studied in a group of hearing-impaired listeners with perceptive high-frequency losses. Individual difference scaling (INDSCAL) and sequential information analysis (SINFA) is applied to a set of consonant-vowel-consonant responses measured under various conditions of speech processing. The processing consists of high-frequency phonemic compression combined with compensation for upward-spread-of-masking (i.e. anti-USOM).In quiet we find an improved perception of frication with compression whereas additional anti-USOM improves the perception of the second and third vowel-formants. In background noise we find remarkably negative effects of anti-USOM on the perception of features containing low-frequency cues, like voicing and nasality. We conclude that the combined results of SINFA and INDSCAL reveal important insight into the possibilities and limitations of phoneme compression and anti-USOM.

5.1 Introduction

Speech processing has become an important issue in recent hearing-aid developments since the introduction of digital techniques. Various kinds of speech-processing schemes are available nowadays that should improve speech understanding in hearing-aid users under different acoustical conditions. However, the actual benefit of most types of processing can not be objectified by speech-intelligibility tests. A limitation of most tests is that the performance with processing is usually expressed in terms of speech scores or speech-reception thresholds. These measures tell us *whether* speech intelligibility improves or not, but they do not show *how* the processing influences the perception of relevant speech cues. Analysis of perceptual speech features may help to identify these fundamental changes in perception that are not shown by speech-intelligibility measures. It can therefore be used to increase the insight in the potentials of a speech-processing scheme. This paper makes this type of analysis for a two-channel speech-processing system and tries to show its validity for optimising signal-processing procedures.

Several methods are available to identify differences in phoneme perception. Miller and Nicely (1955) showed that speech intelligibility can be described by the information that is transmitted by means of articulatory features. Wang and Bilger (1973) used this study as the basis for developing a method called Sequential INFormation Analysis (SINFA). Up to now, only few researchers have used this method to show differences in speech perception due to speech processing (Bustamente and Braida 1987, Dillier and Spillman 1988, van Harten-de Bruijn HE 1997, Wesselkamp and Kollmeier, Yund and Buckles 1995).

A second appropriate tool to distinguish between the perception of speech sounds is INDividual-difference SCALing (INDSCAL). The method was developed by Caroll and Chang (1970) to show differences in speech perception between individual listeners. Bosman (1989) showed that INDSCAL analysis is a very appropriate method to describe differences in listening strategies of different groups of hearing-impaired listeners. Unfortunately, the method was only occasionally used to show differences in performance with various kinds of speech processing (Dillier and Spillman 1988, Dreschler 1988, Frack et al 1999). The few studies that applied INDSCAL or SINFA for this aim usually found small differences that could not be tested on statistical significance. Furthermore, in most cases only one of the two methods was used, missing the opportunity to combine the specific advantages of both methods.

In the present study we exploit both SINFA and INDSCAL to identify differences in phonemic perception due to speech processing. In a previous experiment we found that hearing-impaired listeners benefited from speech-processing based on high-frequency phonemic compression and anti-USOM in conditions *without* background noise (Goedegebure et al 2001). Surprisingly, mainly negative effects were found in conditions *with* background noise. More insight in the perceptual changes caused by the various speech-processing schemes could help us to explain these unexpected results.

The main goal of the present study is to identify the effects of both high-frequency phonemic compression and anti-USOM on perceptual phoneme confusions using INDSCAL and SINFA. The potentials of both methods to analyse effects of speech processing will be considered in addition.

5.2 Materials and Methods

INDSCAL

INDSCAL (Caroll and Chang 1970) is a multi-dimensional scaling method that transforms a set of individual response matrices into two linked n-dimensional spaces. Given the number of dimensions n, INDSCAL calculates an orthogonal set of n perceptual dimensions that explains the maximum amount of variance in the responses. These dimensions constitute a so-called “object space”. Differences between the individual response matrices are expressed as

differences in weights on the perceptual dimensions of the object space. These weight factors are projected in the so-called “subject space”.
In our case, the response matrices contain the phoneme confusions made by individual listeners in different processing conditions. Each phoneme-confusion matrix is transformed into symmetrical similarity matrices using the algorithm suggested by Houtgast and described by Klein et al (1970). This symmetrisation technique is appropriate for our purpose because the applied geometry gives a higher weight to small differences as compared to more conventional techniques like the Euclidean distance. From these symmetrised matrices, INDSCAL calculates an object space in which the phonemes are projected. The distances between the phonemes in this space are related to the amount of confusion: the phonemes cluster when they are confused easily and they are more distant when they are not mixed up. As stated above, each object space is directly linked to a subject space. This name is used as the space usually visualises differences in phoneme confusions between *subjects*. In our way of analysis we mainly used it to show differences in response between *signal-processing conditions.* In that case, the importance of the separate dimensions in the object space is projected for each of the processing conditions. The subject space also shows how well the matching object space fits for each individual matrix. A good fit results in a position close to the unit circle. The Euclidean distance of a vector in the subject space can thus be used as a measure for "goodness-of-fit".
We will mainly show results of a two-dimensional scaling, as higher dimensions did generally lead to a more complex interpretation of the data while the amount of explained variance was not significantly increased. An advantage of 2-dimensional scaling is the possibility of representing the data in simple two-dimensional plots.

SINFA

SINFA (Miller and Nicely 1955) computes per predefined feature (such as voicing, plosiveness, etc.) how much of the input information actually is transmitted. The amount of transmitted information shows the importance of the feature for the perception. A change in the amount of information transmitted by a certain feature indicates that the detection of that feature has changed. The features are defined precisely on phonetic theory before the analysis is done; this is in contrast to the dimensions of the INDSCAL which are calculated and must be interpreted afterwards. SINFA generates sub-matrices per phonetic feature and then computes the amount of transmitted information in bits. The transmitted information is also expressed as a percentage of the input information of that feature. Several iteration steps can be performed to remove redundancy between the transmission rates of the phonetic features. After each iteration step, SINFA fixes the feature with the highest portion of transmitted information. The confusions explained by the fixed features are removed from the analysis in further iteration steps. The rank order in which the features are fixed is commonly used as measure of importance. The higher the rank order of a feature, the more important that feature is for transmitting the phonemic information to the listener.
For quantitative analysis, however, the iteration procedure introduces an undesired order effect. After fixing certain features, only a limited number of confusions are left to determine the transmission rates of the remaining features. Suppose that a feature is the most important

feature in one condition and the third important feature in a second condition. After three iterations in the second condition, the transmitted information for the feature is obtained from a much smaller sub-set of stimuli than obtained after one iteration in the first condition. This leads to an unfair comparison of the transmission rates between both conditions. Therefore, we decided to use the transmission rates of the *first iteration step* for quantitative analysis. This choice implies that the shown transmission rates of the different features are not independent, in contrast to the perceptual dimensions of INDSCAL.

The phonetic features are defined by the feature matrices as shown by tables 5-6a, 5-6b and 5-6c in the Appendix. The phonetic signs are explained in table 5-5. One has to realise that testing was done in the Dutch language. As far as possible a British English counterpart of the sound is given. A major phonetic difference between the Dutch and English is that in the Dutch language a final plosive or fricative is never voiced. The feature matrices are based on a combination of two classifications of the Dutch phonemes [14, 15], taking into account the specific pronunciation of the speaker of the CVC-material. The SINFA-analysis was performed by using the computer program 'FIX', that has been developed at the department Phonetics and Linguistics of UCL, UK.

Relating INDSCAL to SINFA

We related the INDSCAL results to SINFA by using the phonetic features as defined in SINFA to describe the dimensions of INDSCAL. An additional statistical analysis was performed (MANOVA) with the two-dimensional weights of the INDSCAL object space as dependent variables and articulatory features of SINFA as independent factors (see tables 5-5a to 5-5c in the Appendix). We determined the significant interactions between the articulatory features and the INDSCAL dimensions (p-value <0.05). In this way a list of articulatory features could be obtained for each INDSCAL dimension explaining a significant amount of variance. The advantage of this approach is that the interpretation of the INDSCAL dimensions can be objectified and expressed by the same features as used for SINFA.

Experimental conditions

We used the data of a previous experiment (Goedegebure et al 2001) in which the effects of speech-processing have been tested in a group of 14 hearing-impaired listeners for analysis. The implementation of the speech processing conditions and the experimental design are described in detail by Goedegebure et al (2001). We will only give a brief description here.

The signal processing consisted of phonemic compression in a high-frequency channel and linear processing in the low frequencies. The compression was controlled either by the wide-band signal (Wide-Band Control = WBC) or by the filtered speech signal with high-frequency emphasis (High-Frequency Emphasis Control = HFEC). A third configuration consisted of low-frequency reduction in the linear channel to compensate for upward-spread-of-masking effects (Anti-USOM = AU). Note that this anti-USOM low-frequency reduction was performed *in addition* to the wide-band controlled compression within the high-frequency band. The amount of compression was varied in each configuration resulting in a compression ratio (CR) of 1, 2 or 4. This resulted in the eight conditions listed in table 5-1.

Table 5-1 *Definition of the processing conditions by the two dimensions Compression Ratio (CR) and type of processing.*

Type of processing	*CR=1*	*CR=2*	*CR=4*
Wide-Band Control	WBC1	WBC2	WBC4
Anti-USOM	AU1	AU2	AU4
High-frequency Emphasis control	-	HFEC2	HFEC4

We presented our real-time processed stimuli to 14 hearing-impaired subjects. The subjects were selected from a clinical database containing results of pure-tone audiometry and speech audiometry. We selected persons with perceptive moderate–to-severe high-frequency losses and with maximum speech discrimination scores that did not exceed 95%. We obtained speech scores in quiet at the level of maximum performance and in multi-talker background noise at a comfortable presentation level with a signal-to-noise ratio individually fixed near the 50%-score level. Four test sessions were performed. In quiet, we measured the speech score for each condition twice and averaged the results of both measurements.

Speech material

The speech-intelligibility tests were performed with lists of non-sense Consonant-Vowel-Consonant (CVC) words. Each list consisted of 51 CVC words (logatoms) embedded in five different carrier phrases (Steeneken et al 1990). The test words were random combinations of initial consonants, vowels and final consonants, resulting in both nonsense and meaningful words. Successive test words with two or more phonemes in common were rejected. In total 24 different lists were used, spoken by the same male speaker. The carrier phrase was presented on a screen with an open space in which the CVC-word fitted. The subjects were instructed to repeat the word literally, even if it did not make sense, or if he/she only heard part of the word. They were asked to check the answer, after the experimenter had typed it into the computer.

The responses of the subjects were recorded in a confusion matrix. The diagonal of this matrix contains the number of correct responses for each of the stimuli; the other matrix points contain the number of responses for each possible combination of stimulus and incorrect response. The confusion matrices of the initial consonant, the vowel and the final consonant were analysed separately. A different structure of the confusion matrices was required for the two methods of analysis. Responses that were no part of the set of stimuli were left out of the matrix for INDSCAL, as the method requires symmetrical input matrices. With the SINFA-analysis the ‘unknown responses’ were taken in account by defining an extra column with zero's in the feature matrices (not shown in tables 5-5a to 5-5c).

Method of analysis

Two approaches are possible to perform the phonemic analysis. The first approach is to average the confusion matrices *before* applying the analysis, which has the advantage that the phonemic analysis is based on a relatively high amount of phonemic confusions. The second approach is to use the individually obtained confusion matrices as input to the analysis and

apply the averaging afterwards. In that case the significance of within-subject effects can be statistically tested.
INDSCAL based on individual response matrices did not result in reliable or meaningful outcomes, probably because the individual matrices contained a too small number of confusions. Therefore INDSCAL was performed on the confusion matrices averaged per speech-processing condition across subjects as well as on the matrices averaged per subject across speech-processing conditions. The first analysis reveals differences between speech processing conditions and the second analysis reveals differences between subjects.
SINFA could be performed reliably on a smaller amount of confusions than INDSCAL as the method assembles different phoneme-confusions into groups defined by the articulatory features. The individual confusion matrices contained a sufficient number of confusions for a reliable analysis with SINFA as averaging before or after the analysis resulted in very similar quantitative effects. We therefore decided to use SINFA based on the individual response matrices using the advantage of applying statistical tests on individual results. We used different methods for testing statistical significance. For phoneme perception in quiet we chose a non-parametric test as the presence of a ceiling effect violated the assumption of a normal distribution. The effect of compression was evaluated with paired tests between the compression conditions and the corresponding linear conditions (condition AU1 for the anti-USOM programs and condition WBC1 for the compression conditions without anti-USOM). Next, the effect of type of processing was tested by comparing the AU- and HFEC-conditions with the corresponding WBC-conditions. In background noise we applied a repeated-measures ANOVA strategy with “CR” and “type of processing” as within-subject factors. SPSS-9.0 was used for all statistical testing.

5.3 Results

INDSCAL: fit of individual listeners

First, we performed INDSCAL on the response matrices averaged *per listener* across different signal-processing conditions to investigate if the listeners used a comparable perceptual strategy. In quiet, a good fit was obtained for all listeners. For each of the listeners in each of the three phoneme conditions we obtained a goodness-of-fit higher than 0.8. It shows that the perception of the individual listeners is described well by a common two-dimensional perceptual space. The individual loads on each of the perceptual dimensions did not significantly deviate from a normal distribution (Kolmogorov-Smirnov $p>0.05$) and outliers were only found incidentally.
In background noise we also found a good fit for all listeners except for one. This listener was characterised by a relatively poor fit for both initial- and final consonant perception (a goodness-of-fit of 0.66 and 0.76 respectively), corresponding to deviating individual loads. The confusion matrices of this subject contained many unknown responses, especially for the conditions with anti-USOM. Apparently, the very poor performance with anti-USOM considerably influences the results of INDSCAL for this listener. We repeated INDSCAL on the group data without the confusions of this listener, but no relevant differences were found.

We conclude that the perception of the individual listeners is generally described well by a common two-dimensional perceptual space. Therefore we will use the response matrices averaged across all listeners henceforth.

INDSCAL: analysis of perceptual dimensions

The results of INDSCAL for the matrices averaged per speech-processing condition are presented in figures 5-1a to 5-1d (initial-consonant perception), 5-2a to 5-2d (final-consonant perception) and 5-3a to 5-3d (vowel perception). The a- and c-figures contain the object spaces in quiet and background noise respectively, whereas the b- and d-figures contain the linked subject spaces. The interpretation of the INDSCAL dimensions is objectified in table 5-2. This table lists the articulatory features explaining a significant amount of variance (MANOVA, $p<0.05$) of each perceptual INDSCAL dimension, as well as the total amount of variance explained by each of the dimensions. First we will discuss the interpretation of the various perceptual dimensions. Then the effects of speech processing will be considered.

Table 5-2 *Articulatory features explaining a significant amount of variance (MANOVA, p<0.05) of the perceptual dimensions as defined by INDSCAL.*

phoneme	***quiet***				***noise***			
	dimension I		*dimension II*		*dimension I*		*dimension II*	
category	*features*	*var*	*features*	*var*	*features*	*var*	*features*	*var*
C1	fri, plo*	0.53	nas, voi	0.34	nas, voi, plo	0.49	fri	0.37
C2	nas, voi, plo	0.61	fri	0.31	nas, voi	0.65	plo	0.18
V	F1	0.63	F2, F3*	0.26	F1	0.67	F2,F3*	0.20

INDSCAL: consonant perception

Figure 5-1a shows the INDSCAL object space for initial-consonant perception *in quiet*. A clustering of consonants is found according to the articulatory features frication, plosiveness and nasality (encircled by the dotted lines). The fricatives are found in the upper *left* corner and the plosives in the upper *right* corner, which shows that the first dimension is mainly defined by frication and plosiveness. The second dimension is mainly defined by nasality as the nasals cluster below in the figure. It is also related to voicing as the voiced consonants are always positioned beneath the unvoiced ones. In the middle of the object space we find the glides and the liquids. Table 5-2 confirms that the first dimension is related to frication and plosiveness whereas the second dimension is related to voicing and nasality.

A different situation is found for initial-consonant perception *in background noise*, as shown in figure 5-1c. The first dimension in figure 5-1c is similar to the second dimension of figure 5-1a as it is mainly defined by nasality. However, the second dimension of figure 5-1c is a kind of simplified version of the first dimension of figure 5-1a. The fricative-plosive distinction has almost disappeared except of the distinction of /s/ and /z/ (sibilance) from the other consonants. The other fricatives are shifted to the cluster with plosives (indicated by the arrow), which means that they can hardly be distinguished from the plosives. Table 5-2 confirms that the other three important features nasality, voice and plosiveness constitute the first dimension and thus explain most of the variance in background noise.

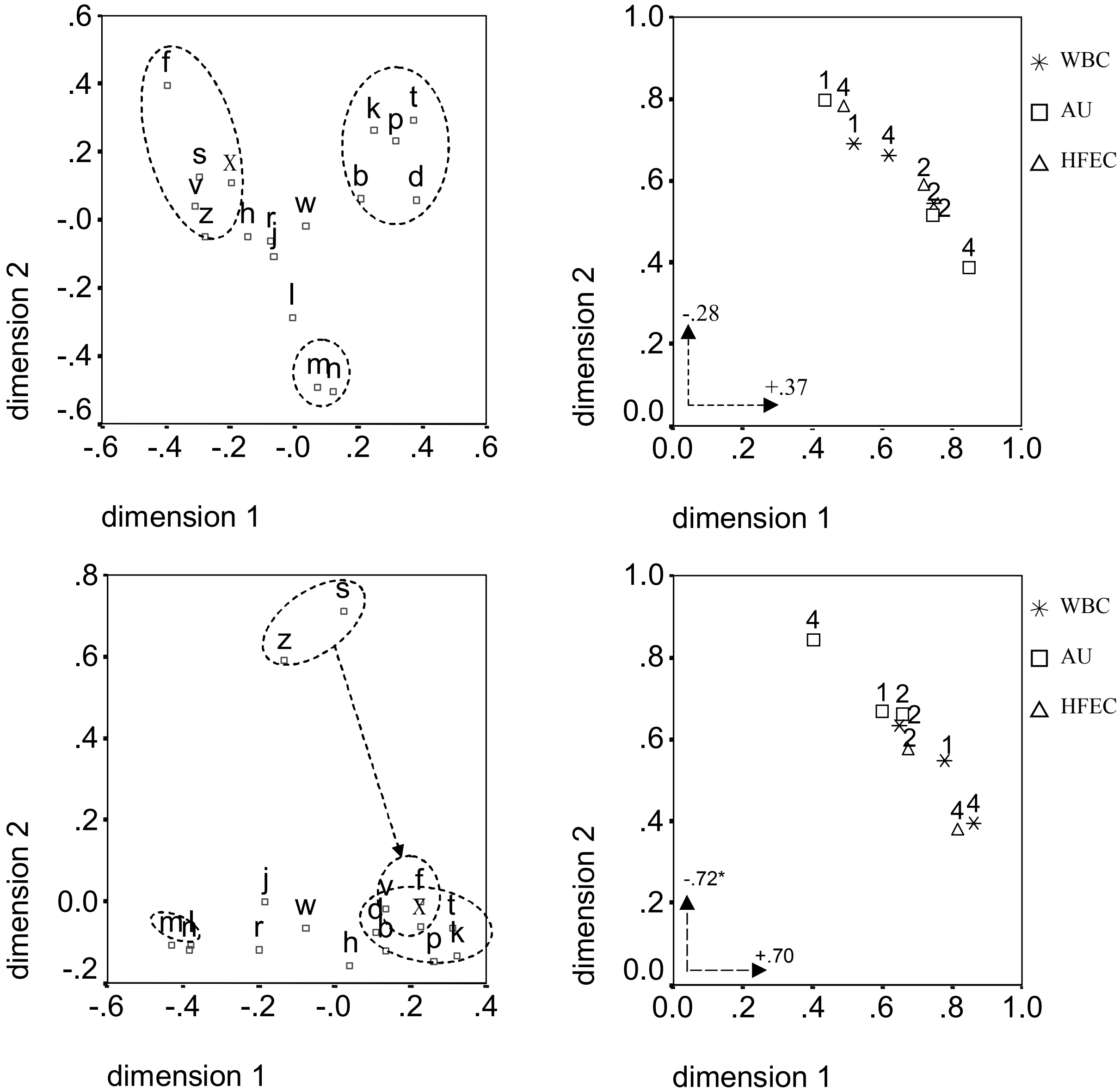


Figure 5-1a/b/c/d *INDSCAL results for initial-consonant confusions averaged per processing condition across subjects. Figures 1a and 1c show the object spaces obtained in quiet and in background noise respectively. The clusters with nasals, fricatives and plosives are encircled. The cluster with the arrow in figure 1c shows the group of fricatives that it is shifted to the corner with plosives due to the presence of background noise. Figures 1b and 1d show the loads of the response matrices on the linked both object spaces (figures 1a and 1c respectively). The markers in figures 1b and 1d are categorised by type of configuration (stars=WBC, squares=AU, triangles=HFEC), while the labels represent the compression ratio used. In the corner left below of figures 1b and 1d the correlation is shown between the loads on the dimension and the phoneme score. Significant correlations are indicated by an asterix (p<0.05).*

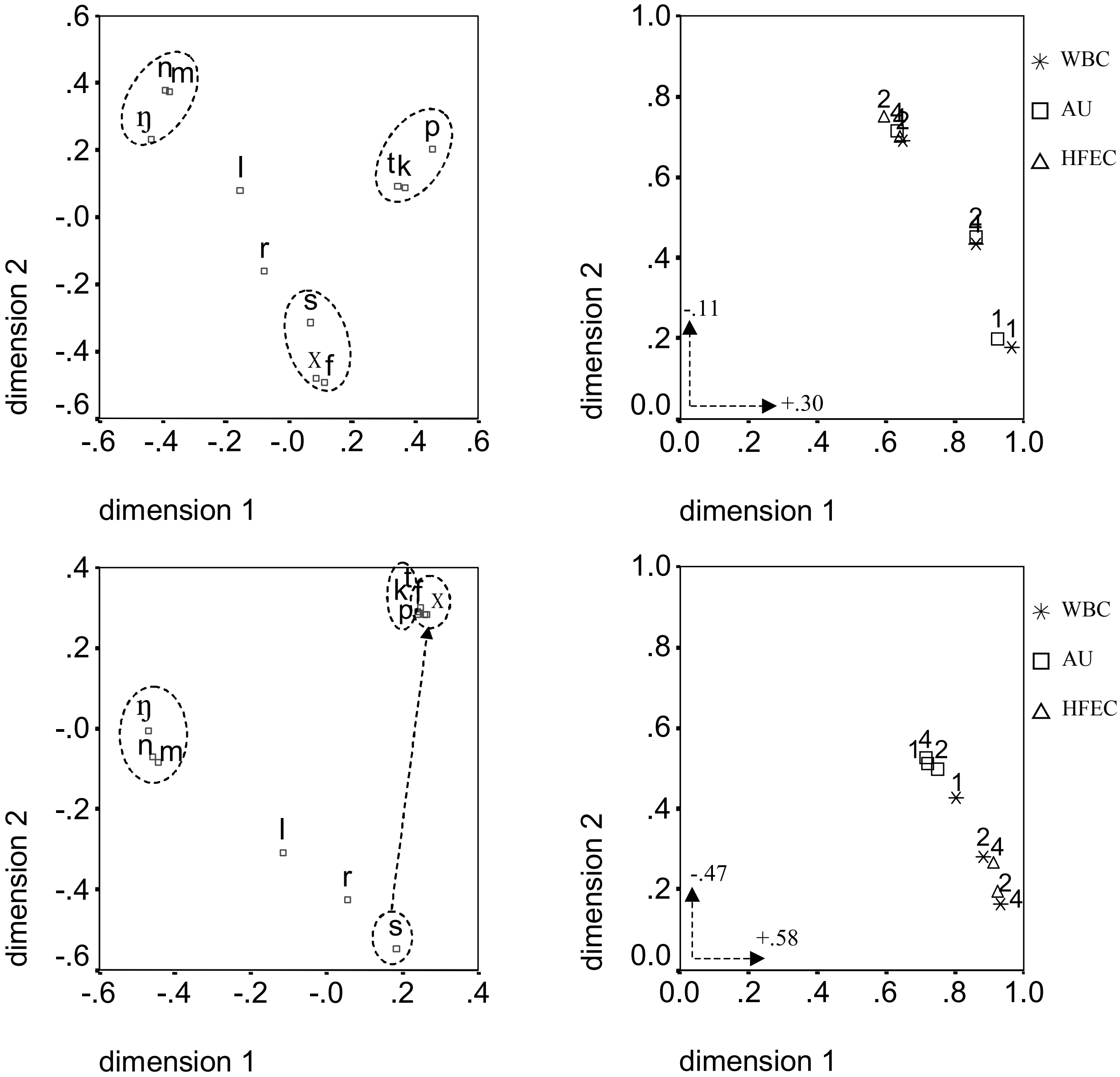


Figure 5-2a/b/c/d *INDSCAL results for final-consonant confusions averaged per processing condition across subjects. Figures 1a and 1c show the object spaces obtained in quiet and in background noise respectively. The clusters with nasals, fricatives and plosives are encircled. The cluster with the arrow in figure 1c shows the group of fricatives that it is shifted to the corner with plosives due to the presence of background noise. Figures 1b and 1d show the loads of the response matrices on the linked both object spaces (figures 1a and 1c respectively). The markers in figures 1b and 1d are categorised by type of configuration (stars=WBC, squares=AU, triangles=HFEC), while the labels represent the compression ratio used. In the corner left below of figures 1b and 1d the correlation is shown between the loads on the dimension and the phoneme score. Significant correlations are indicated by an asterix (p<0.05).*

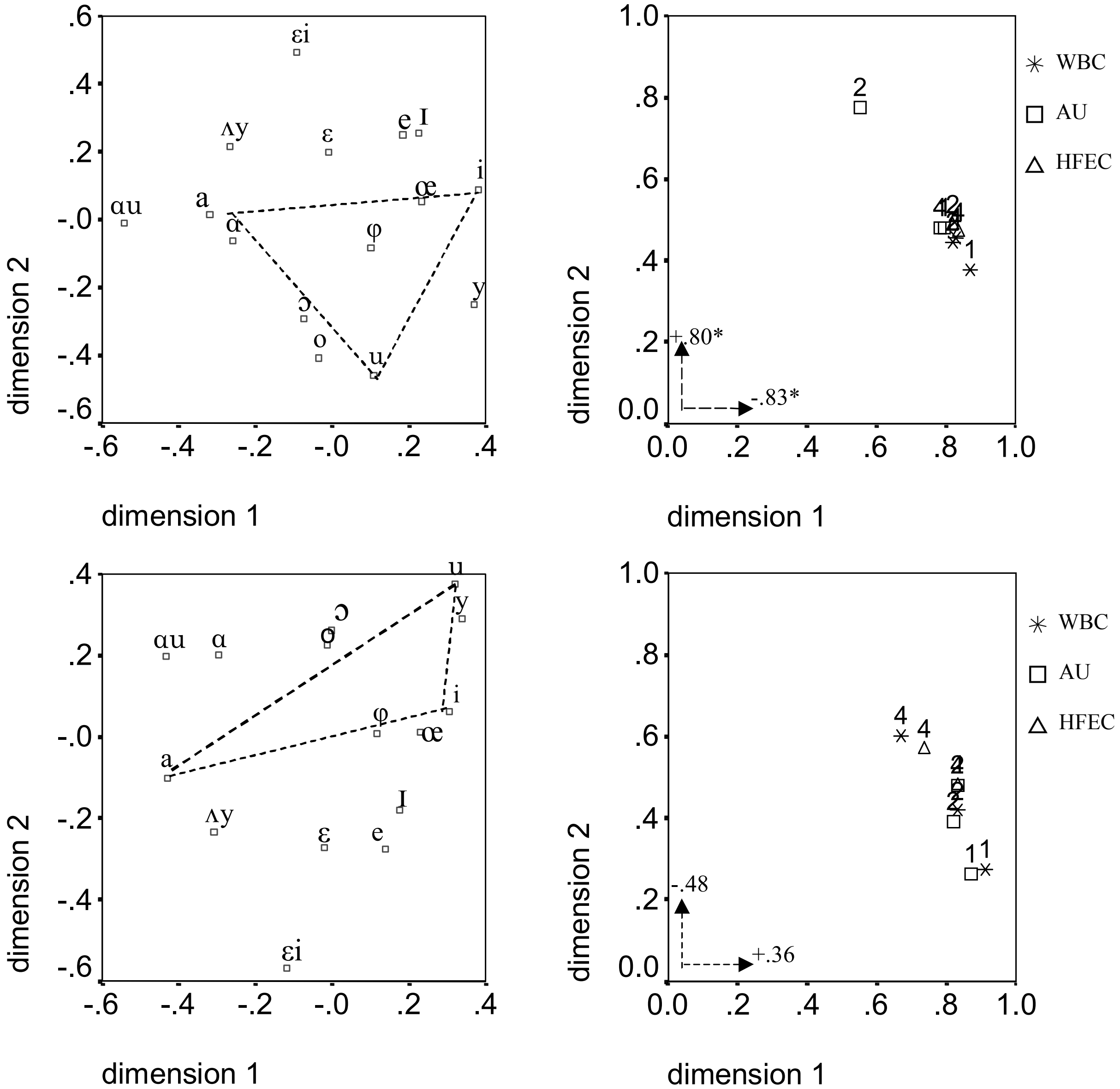


Figure 5-3a/b/c/d *INDSCAL results for vowel confusions averaged per processing condition across subjects. Figures 1a and 1c show the object spaces obtained in quiet and in background noise respectively. The triangles constituted by the first two formants is drawn in both figures 1a and 1c. The markers in figures 1b and 1d are categorised by type of configuration (stars=WBC, squares=AU, triangles=HFEC), while the labels represent the compression ratio used. In the corner left below of figures 1b and 1d the correlation is shown between the loads on the dimension and the phoneme score. Significant correlations are indicated by an asterix (p<0.05).*

Figures 5-2a and 5-2c show that a very similar result is obtained for final-consonant perception. The object space is mainly constituted by the three clusters as defined by frication, plosiveness and nasality. In background noise, a part of the cluster with fricatives moves to the plosive-cluster (indicated by the arrow). This means that again plosive-fricative distinction is seriously degraded due to background-noise masking.

INDSCAL: vowel perception

For vowel understanding in quiet the perceptual dimensions are dominated by the three formants. The first two formants are projected on separate dimensions; the first formant on the first dimension and the second formant on the second dimension. The triangle constituted by both formants is drawn in figures 5-3a and 5-3c. In quiet the triangle has a nice symmetrical appearance indicating that both first and second formant significantly contribute to vowel perception. In background noise the triangle is a bit squashed, mainly because the distance is reduced between vowels that significantly differ in second-formant frequency (such as o-a, ɔ- ɑ, u- i). It seems that the background-noise masking has a more disturbing effect on second-formant than on first-formant perception.

Less obvious is the role of the third formant as it explains a significant amount of variance of both dimensions (see table 5-2). Three-dimensional INDSCAL did not separate the third formant from the other two formants. The dependency with the first two formants is probably too high to be projected on a separate orthogonal dimension. We therefore continued with the results of the two-dimensional scaling. When inspecting figures 5-3a and 5-3c carefully, the third formant seems to be related more to the second dimension than to the first dimension. The second dimension distinguishes better between pairs of vowels that mainly differ in third-formant frequency, such as a -ɑ, o-ɔ and y-œ. Therefore we will consider the second dimension of vowel perception to be related to both second- and third-formant perception.

INDSCAL: effects of speech processing

The effects of speech processing can be obtained from the loads of the response matrices on the linked perceptual dimensions as shown in figures 5-1b, 5-1d, 5-2b, 5-2c, 5-3b and 5-3d. Each of the figures is linked to the preceding object space. The markers in both figures are categorised by type of configuration (stars=WBC, squares=AU, triangles =HFEC), while the labels represent the compression ratio used. The impact of the loads on phoneme intelligibility is indicated by the correlation coefficient with the phoneme score per processing condition in the corner left below. Significant correlations are indicated by an asterix. A positive correlation means that a higher weight on that perceptual dimension corresponds to an improved phoneme intelligibility.

Figure 5-1b shows that the conditions with linear processing (CR=1) mainly weigh on the second dimension (nasality and voicing), whereas the compression conditions mainly cluster on the first dimension (frication and plosiveness). This might be interpreted as an improved identification of frication and plosiveness due to phoneme compression. The positive correlation between the weighing on the first dimension and phoneme score seems to support the given interpretation, but it should be noted that the correlation is weak and statistically non-significant ($p>0.05$).

A distinction of compression and linear conditions is also found for the final-consonant perception in quiet (see figure 5-2b). The linear conditions mainly weigh on the first dimension whereas the compression conditions have also a substantial weight on the second dimension which is related to frication. Again, it seems that compression improves the distinction of frication. However, the correlation of the phoneme score and the second-dimension weight is almost zero (-0.11) so in this case the difference in phoneme perception does not lead to an improved speech intelligibility.

In background noise, the use of anti-USOM mainly determines the effects of processing on both initial- and final-consonant perception. Figures 5-1d and 5-2d show that the conditions with anti-USOM (squares) cluster with a relatively high weighing on the second dimension (frication) and a low weighing on the first dimension (nasality, voicing, plosiveness). The impact of this clustering on speech intelligibility is negative, as the first dimension is positively correlated with phoneme score. Apparently, anti-USOM degrades the identification of the three important features nasality, voicing and plosiveness in background noise. Obviously, only limited compensation is obtained by a higher load on the second dimension as it only explains the distinction of the two sibilants /s/ and /z/ from the other consonants.

Figure 5-3b shows that for vowel perception in quiet combined compression with CR=2 and anti-USOM (condition AU2) is separated from the other conditions. Compared to the other processing conditions, this condition weighs mainly on the second dimension related which is to the second and third formant. The positive correlation of the loads on the second dimension and the vowel score indicates that more second- and third-formant information is made available by compensation of anti-USOM in combination with CR=2. This means that the anti-USOM does what it should do, at least in combination with high-frequency compression.

In background noise no such positive effect of anti-USOM is found. Instead, a difference is found between the linear conditions and all compression conditions (see figures 5-3d). The compression conditions weigh generally less on the first dimension, which has a negative effect on the phoneme score. Compression seems to degrade the perception of first-formant information in background noise. This is a bit surprising as the compression is mainly active at high frequencies, whereas the first formant is usually found at relatively low frequencies. Somehow the use of compression affects certain vowel cues that are (also) related to the first dimension.

SINFA

A major difference between INDSCAL and SINFA is that SINFA gives us the information transmitted on predefined articulatory features allowing for an easy interpretation of the data. The significant effects on the amount of transmitted information are presented in figures 5-4 to 5-6. The transmission rates are shown as function of compression ratio whereas the different types of processing are indicated by different markers (stars for WBC, squares for AU and triangles for HFEC). Significant differences with the linear reference are indicated by an “S” (Wilcoxon, $p<0.05$).

SINFA: consonant perception quiet

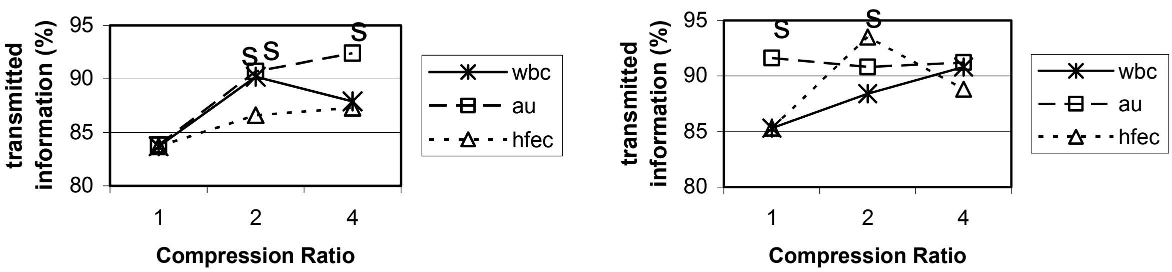


Figure 5-4a/b *Percentages of initial-consonant information transmitted* ***in quiet*** *by the features* ***frication****(a) and* ***nasality*** *(b) respectively as function of compression ratio. The three types of compression are represented by different lines and markers (WBC= continuous line and stars, AU= dashed line and squares, HFEC=dotted line and triangles). Significant effects compared to linear processing are indicated by an S (Wilcoxon, p<0.05).*

For the initial consonants in quiet, the main improvements with compression are found for the information transmitted by frication. Compared to linear processing, the improvement is statistically significant for frication with conditions WBC2, AU2 and AU4 (Wilcoxon, $p<0.05$). Figure 5-4a illustrates the increase in transmitted information as function of the compression ratio. For the three mentioned conditions the improvement is more than 5% on average. Note that only a slight improvement is found for the conditions with high-frequency emphasis control (triangles), which shows that the use of wide-band control is mainly responsible for the improved transmission of frication. A similar consistent improvement with compression is found for the transmission of plosiveness, but the effect is not statistically significant (Wilcoxon, $p>0.05$).

Another significant improvement on initial-consonant perception is found for the transmission of nasality (see figure 5-4b). Compared to linear processing, a significant increase is found for conditions HFEC2 and AU1 (Wilcoxon, $p<0.05$). The fact that an improvement is found for a condition with only anti-USOM and *no* compression (AU1) suggests that it is mainly caused by compensation of spectral-masking effects.

With compression, the transmission rates for the *final*-consonant features plosiveness and frication improve as well. However, none of the found improvements are significant (Wilcoxon, $p>0.05$). This is not surprising as a ceiling effect is found for the information transmitted with linear processing. For the four most important final-consonant features, including frication and plosiveness, the *median* value is 100%! This means that most subjects do not profit from compression in case of final-consonant perception because they perceive already the maximum amount of articulatory information with linear processing. The majority of final-consonant confusions were caused by a disturbed perception of the feature “place”. Unfortunately, neither compression nor anti-USOM increases the amount of information transmitted by this feature.

SINFA: consonant perception in background noise

Table 5-3 *Summary of repeated-measures ANOVA on the information transmitted in background noise by consonant features with within-subject factors type of processing and CR (not shown in the table).*

feature	*processing C1*	*constrasts (to WBC)*		*processing C2*	*constrasts (to WBC)*	
		AU	*HFEC*		*AU*	*HFEC*
voi	F(2,12) =21.90**	-**	n.s.	F(2,12)=6.62*	-**	n.s.
nas	F(2,12) =16.54**	-**	n.s.	F(2,12)=19.40**	-**	n.s.
plo	F(2,12) = 5.18 *	-**	n.s.	n.s.	n.t.	n.t.
fri	n.s	n.t.	n.t.	n.s.	n.t.	n.t.
pla	F(2,12) = 4.57 *	-**	n.s.	n.s.	n.t.	n.t.

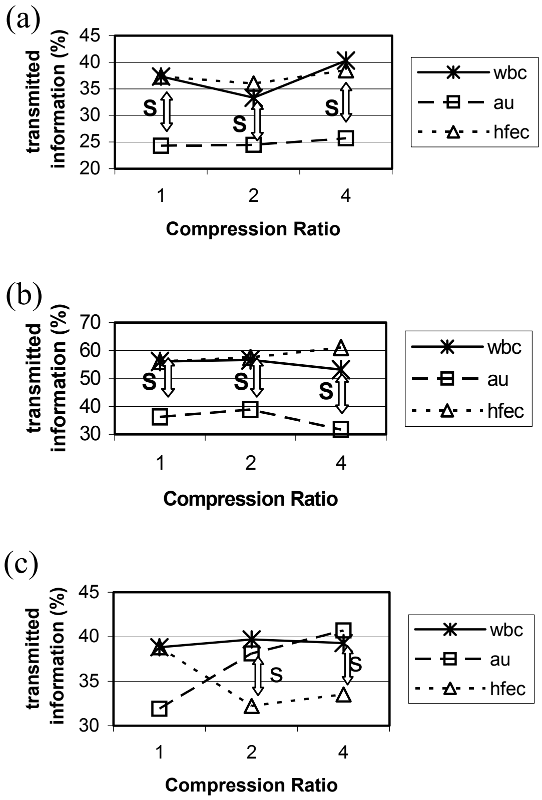


Figure 5-5a/b/c *Percentages of initial-consonant information transmitted in* ***background noise*** *by the features* ***voicing*** *(a),* ***nasality*** *(b) and* ***frication*** *(c) respectively as function of compression ratio. The three types of compression are represented by different lines and markers (WBC= continuous line and stars, AU= dashed line and squares, HFEC=dotted line and triangles). Significant contrasts between different types of compression are indicated by an S (Repeated Measures ANOVA, p<0.05)*

In background noise, SINFA confirms the result of INDSCAL that anti-USOM processing has a negative effect on consonant perception. The results of a repeated-measures ANOVA with "CR" and "type of processing" are found in table 5-3. Significant negative effects of anti-USOM processing are found for the features voicing and nasality (both initial- and final consonant, repeated measures ANOVA $p<0.01$), as well as for the initial-consonant features plosiveness and place (repeated measures ANOVA $p<0.05$). Figures 5-5a and 5-5b illustrate the large difference in information transmitted by voicing and nasality respectively between the anti-USOM conditions (squares) and the other conditions (triangles and stars). The poorer transmission is obviously caused by the anti-USOM processing and *not* by the amount of

compression. The amount of transmitted information does not depend on compression ratio at all. Frication is the only feature of which the perception remains relatively unaffected by anti-USOM processing (see figure 5-5c). This confirms the relatively high load of anti-USOM conditions on the second INDSCAL dimension in figure 5-3b. Similar results are found for the final-consonant perception in background noise, with again significant negative effects of anti-USOM on the features voicing and nasality (repeated measures ANOVA, $p<0.01$).
Next to the described negative effects of anti-USOM, another significant negative effect is found for High-Frequency Emphasis Control (HFEC) on the initial-consonant transmission of frication (see figure 5-5c, repeated measures ANOVA, $p<0.05$). A reason for this may be the reduced emphasis of consonant cues in the high-frequency band compared to wide-band controlled compression.

SINFA: vowel perception in quiet

SINFA of vowel confusions in quiet revealed two interesting improvements. For configuration AU2 the amount of information transmitted by both the second and the third formant increases significantly (Wilcoxon, $p<0.05$). The effect is illustrated for the second formant in figure 5-6a. *No* statistical improvement is found for the first formant, which indicates that anti-USOM mainly enhances second- and third formant information. Apparently, the anti-USOM filter functions optimally in combination with CR=2 as none of the other configurations revealed a significant improvement for any of the vowel features. This is in agreement with the distinctively higher load of condition AU2 on the second INDSCAL dimension.

SINFA: vowel perception in background noise

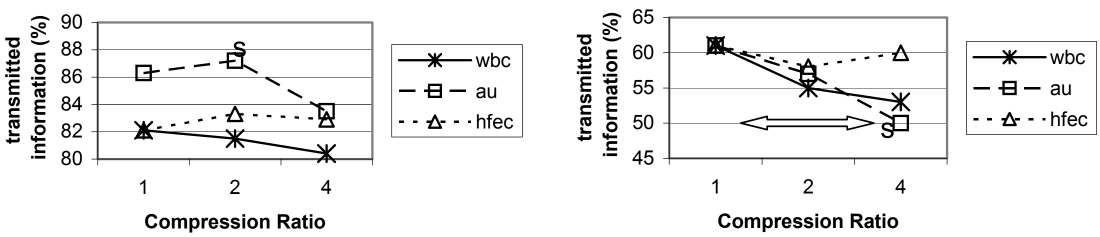


Figure 5-6a/b *Percentages of vowel information transmitted by the second formant in quiet and in background noise respectively as function of compression ratio. The three types of compression are represented by different lines and markers (WBC= continuous line and stars, AU= dashed line and squares, HFEC=dotted line and triangles). In figure 5-6 a the S indicates a significant effect compared to linear processing (Wilcoxon, p<0.05), while in figure5-6 b the S indicates a significant contrast between CR=4 and CR=1 (Repeated Measures ANOVA, p<0.05)*

For vowel perception in background noise the transmission rates depend of the amount of compression instead of the use of anti-USOM processing. The results of a repeated-measures ANOVA with "CR" and "type of processing" are found in table 5-4.

feature	*CR Vowel*	*contrasts (to CR = 1)*	
		CR=2	*CR=4*
F1	F(2,12) =13.22**	n.s.	-**
F2	F(2,12) = 3.99*	n.s.	-*
F3	F(2,12) = 12.30**	n.s.	-**
dip	n.s.	n.t	n.t
dur	n.s.	n.t	n.t

Table 5-4 *Summary of repeated-measures ANOVA on the information transmitted in background noise by vowel features with within-subject factors type of processing (not shown in this table) and CR.*

Significant negative effects of compression with CR=4 are found for each of the three formants (repeated measures ANOVA, $p<0.05$). It seems that a high amount of phonemic compression affects the identification of all relevant vowel features in background noise. The effect is illustrated for the second formant in figure 5-6b. The amount of transmitted information obviously decreases as function of compression ratio, except for the compression conditions with high-frequency emphasis. Interestingly, the compression conditions with high-frequency emphasis control do not show the large negative effects on vowel perception found for the other compression conditions. Even a significant positive effect is found for these conditions on the perception of the feature "duration" (main effect: $F(2,12)=12.47$, $p<0.01$). It seems that the high-frequency emphasis filtering in the control signal effectively compensates for negative effects of compression on vowel perception in background noise.

5.4 Discussion

Combined use of SINFA and INDSCAL

Effects of signal processing on phoneme perception can be adequately identified by combined use of INDSCAL and SINFA. Both methods reveal corresponding, but also complementary effects. INDSCAL identifies and visualises the most relevant differences in phoneme perception. However the interpretation of these differences is not always easy because the perceptual dimensions can be related to several perceptual features. An advantage of the additional use of SINFA is that the various effects are separated for the different predefined features.

We facilitated the integrated use of both methods by matching the feature matrix of SINFA to the object space of INDSCAL. An adequate and objective interpretation of the dimensions was obtained by performing MANOVA analysis with the predefined features as factor. This considerably reduces the subjectivity of the INDSCAL method, as the interpretation of the different perceptual dimensions is sometimes not obvious. A second important methodological improvement was that we could apply statistical tests to effects measured with SINFA. The analysis of the individual response matrices did not lead to noticeable problems and allowed us to apply MANOVA to the individual transmission rates. We strongly recommend the use of statistical testing, as it takes inter-subject differences into account and helps to identify the consistent patterns in the data. In our opinion, the performed adjustments and the combined use of both methods are necessary to interpret the phonemic analysis in a meaningful way.

Effects of speech processing on perception

Both INDSCAL and SINFA are in agreement about three prominent effects:

1. improved transmission of frication in quiet with compression
2. enhanced second- and third formant perception in quiet with combined compression and anti-USOM
3. poorer transmission of voicing and nasality in background noise with anti-USOM

These consistent effects on perception explain the main effects of compression and anti-USOM on speech intelligibility measured by Goedegebure et al (2001). The positive effects of compression on the consonant score in quiet are mainly explained by an improved identification of frication and plosiveness. This is what we expected and why we used compression. Hearing-impaired listeners with high-frequency losses have in general problems with the identification of fricatives and plosives as the information of these features is mainly transmitted by temporal cues and at high frequencies. The fast-acting compression system improves the detection of these features by emphasising the relatively weak speech cues within the high-frequency band in the temporal domain. With an attack time of 5 ms and a release time of 15 ms our compression system is fast enough to be effective. Bustamente and Braida (1987) also found that fast-acting compression improves the transmission of weak fricatives and plosives. However, these effects were counteracted in their experiments by negative effects on other consonants. We did not find such counteracting negative effects, probably because our compression system is only active in the high-frequency channel and does not affect information in more low-frequency bands. Another reason may be the use of "overshoot-reduction" in our compression system. The study of Dreschler (1988) reports a deviating result as he found that wide-dynamic-range compression reduces the perceptual weight of the frication dimension. He suggested that overshoots at the beginning of a word might be responsible for this effect. In our system we compensated for overshoots which can explain the more favourable effects of phoneme compression. This indicates that reducing the temporal distortion caused by fast compression may have a positive effect on consonant perception.

Yund and Buckles (1995) report benefits on the perception of high-frequency cues with their multi-channel system that look very similar to our results. This suggests that both single-channel high-frequency and multi-channel compression make more high-frequency information available. Still, our results suggest that the effect is stronger for wide-band control than for high-frequency control. An advantage of wide-band controlled compression is that consonant cues in the high-frequency band are generally emphasised compared to vowel cues, as in the wide-band signal the vowels contain more energy than the consonants. When the high-frequency content dominates the compression control, the consonants containing much high-frequency energy (like fricatives and plosives) will be hardly emphasised or even suppressed. Our results with HFEC compression suggest that this leads to a less effective transmission of these features compared to wide-band controlled compression. With HFEC, the positive effects of compression on the transmission of frication is only limited in quiet (see figure 5-4a) and the effect is even negative in background noise (see figure 5-5c).

The results of the phonetic analyses also explain the limited benefit of compression on final-consonant intelligibility as reported in Goedegebure et al (2001). At first sight, we would expect similar improvements for both initial and final-consonants as the features frication and plosiveness are enhanced in a very similar way. However, a ceiling effect is found for the information transmitted by plosiveness and frication which is not present for the initial consonant. As a consequence, room for improved transmission of frication and plosiveness is only available for a minority of listeners. The feature place is responsible for the majority of final-consonant confusions within the population, but unfortunately compression has no overall effect on the perception of this feature. This is not surprising as the difference in articulation place results in very subtle spectral differences, such as the differences between /n/, /m/ and /ŋ/. Even normal-hearing listeners have difficulties to identify these differences. We might expect more benefits from compression on the final-consonant intelligibility in a group of subjects with poorer speech discrimination, as these subjects will show lower transmission rates for the features frication or plosiveness. Phoneme compression should therefore preferably be used if there are substantial problems with the perception of frication or plosiveness in the aided condition.

The lack of benefits in background noise with compression can be attributed to the difference in availability of high-frequency speech cues. Essential high-frequency consonant cues are masked by the noise. Especially in our group of subjects with moderate-to-severe perceptive high-frequency losses we may expect a distorted perception of high-frequency cues due to the poor cochlear high-frequency processing. Figure 5-3a, for instance, shows that the initial-consonant features frication and plosiveness cannot be clearly distinguished in background noise. Single-channel compression cannot compensate for this masking effect as it does not improve the signal-to-noise ratio within the compression channel. Emphasising high-frequency cues will not have the desired positive effects as long as the background noise is equally emphasised. Therefore, an effective noise-reduction system in front of the compressor is needed to take advantage of high-frequency phoneme compression in background-noise conditions.

The concept of combining anti-USOM and compression works out the way we had expected in conditions without background noise. When using a compression ratio of 2 the masking of the second and third formant by the lower formants is effectively reduced. The applied combination of filtering and compression apparently compensates for USOM caused by the reduced frequency selectivity at middle and high frequencies.
We would expect to find at least a similar positive effect in conditions with background noise as spectral-masking effects should increase due to the noise. The strong negative effect of anti-USOM in background noise is therefore very disappointing and asks for an explanation. As already has been mentioned in the previous paragraph, the availability of high-frequency speech cues is strongly reduced for our subject population in critical signal-to-noise ratios. Our results show that the background noise affects the perception of low-frequency cues less than the perception of high-frequency cues. Most of the speech information is transmitted by

the first formant, voicing and nasality, whereas the distinction between plosiveness and frication is completely reduced to the recognition of sibilance (/z/ and /s/). Similar findings are reported by Miller and Nicely (1955). Additional emphasis of the high-frequencies by anti-USOM filtering or compression does not increase the availability of high-frequency cues. The transmission of frication in background noise, for instance, is hardly influenced by using anti-USOM or compression (see figure 5-5c). Masking effects *within* the high-frequency band seem to play a more important role than the USOM effects that remain after having applied a half-gain frequency shaping. A reduction of low-frequency energy, on the other hand, may have detrimental effects on the performance as INDSCAL shows that the low-frequency cues become more relevant in background noise conditions. The large negative effects found with anti-USOM on the transmission of voicing and nasality indicate that essential low-frequency cues are indeed seriously suppressed. This is at least surprising as anti-USOM is often applied to improve speech perception in background noise. Our results indicate that prudence is called for when applying anti-USOM in critical speech-in-noise conditions. Positive effects of anti-USOM may show up in the case of low-frequency noise or less critical signal-to-noise-ratios but these conditions were not investigated in this study. Finally, the use of high-frequency emphasis control did not result in the remarkable negative effects in background noise found with anti-USOM, most probably because it enhances the weaker higher formants *without* affecting the low-frequency cues.

5.5 Conclusions

- SINFA and INDSCAL reveal useful, consistent and complementary information about the effects of speech processing on phoneme perception. Combining both methods is needed for a clear interpretation of the analysis data.
- Fast-acting single-channel compression within the high-frequency band enhances the distinction of frication and plosiveness in quiet. Benefits are mainly found for the initial-consonant perception as for the final-consonant perception the identification of frication and plosiveness is already performed well with linear processing. At low signal-to-noise ratio the advantage of phoneme compression disappears as the perception of high-frequency features is too much disturbed by the masking effect of the noise.
- Anti-USOM in combination with fast-acting compression improves the transmission of the second and third formant of vowels in quiet. In conditions with background noise, however, the low-frequency reduction seriously disturbs the transmission of features containing low-frequency information, such as voicing and nasality.

5.6 Acknowledgements

This research has been performed in the framework of the HEARDIP-project, which was part of the Technological Innovative programme for the Disabled and Elderly (TIDE) and was funded by the European Commission. We want to direct our special thanks to the department

Phonetics and Linguistics of UCL in London to provide us with their SINFA program 'FIX'. Furthermore, we would like to thank Arjen Bosman that wrote an important part of the INDSCAL software we used. Marjon Droogendijk, Marieke van Capellen en Helene van Harten-de Bruyn are thanked for the fruitful discussions about the phonetic methods used in this paper. We also acknowledge the contributions made by the Heinsius Houbolt Foundation to support this research.

5.7 References

Bosman AJ (1989) Speech perception by the hearing impaired. University of Utrecht: Utrecht (Dissertation)

Bustamente DK, Braida LD (1987) Principal-component amplitude compression for the hearing impaired. J Acoust Soc Am 82:1227-1242.

Caroll JD, Chang J (1970) Analysis of individual differences in multidimensional scaling via N-way generalisation of 'Eckart-Young' decompensation. Psychometrika 35(3):288-319.

Dillier N, Spillmann T (1988) Wahrnemung akustischer Sprachmerkmale mit einkanaligen Cochlea-Implantaten und Hörgeräten. HNO 36(8):335-341.

Dreschler WA (1988) The effects of specific compression settings on phoneme identification in hearing-impaired subjects. Scand Audiol 17:35-43.

Franck BAM, Kreveld-Bos Sv, Dreschler WA, Verschuure J (1999) Evaluation of spectral enhancement in hearing aids, combined with syllabic compression. J Acoust Soc Am; 106 (3):1452-1464.

Goedegebure A, Hulshof M, Verschuure J, Dreschler WA (2001) Effects of Single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners. Audiology 40:10-25.

van Harten-de Bruijn HE, van Kreveld-Bos CSGM, Dreschler WA, Verschuure J (1997) Design of two syllabic non-linear multi-channel signal processors and the results of speech tests in noise. Ear Hear 18:26-33.

Klein W, Plomp R, Pols LCW (1970) Vowel spectra, vowel spaces and vowel identification. J Acoust Soc Am 48:999-1009.

Miller GA, Nicely PA (1955) An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 27(2):338-352.

Nooteboom SG, Cohen A (1988) Spreken en verstaan: een nieuwe inleiding tot de experimentele fonetiek. Third release ed. Assen: Van Gorcum.

Steeneken HJM, Geurtsen FWM, Agterhuis E (1990) Speech data-base for intelligibility and speech quality measurements. TNO-Institute for Perception: Soesterberg.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Wang MD, Bilger RC (1973) Consonant features in noise: a study of perceptual features. J Acoust Soc Am 54(5):1248-1266.

Wesselkamp M, Kollmeier B (1991) Analyse von Vokalverwechslungen an hand akustischer Sprachmerkmale. Fortschritte de Akustik-GA, 91. Bad Honnef: DPG-Kongress-Verlag.

Vieregge WH (1985) Transcriptie van spraak. Dordrecht: Foris Puclication Holland.

Yund EW, Buckles KM (1995) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239.

5.8 Appendix Definition of articulatory features

A phonetic transcription of the phonemes that are part of our CVC-material is listed in table 5-5. Dutch examples of words containing the phoneme are given as well as British English counterparts with a similar pronunciation. The definition of the articulatory features as used for SINFA are given in tables 5-6a to 5-6c. The consonant features listed in tables 5-6a and 5-6b are frication (fri), nasality (nas), plosiveness (plo), voicing (voi) and place (pla). The feature place is defined by the six articulation places **a**lveolair (coronal); **b**ilabial; **l**abiodental; **v**elar; **p**alatal and **u**vular. In table 5-6b the voiced plosives and fricatives are omitted because in Dutch these sounds are pronounced voiceless when they are used in a final position; in certain dialects voicing of most plosives and fricatives is not used.
The vowel features shown in table 5-6c are the three vowel formants (F1, F2 and F3 respectively), diphthong (dip) and duration (dur). The first formant is split into four classifications according to the position of the tongue: 1-low, 2-half low, 3-half high, 4-high. The second formant is divided in the three horizontal tongue-positions: front (f), middle (m) and back (b). The distribution of unrounded (u), rounded (r) and spread vowels (s) is shown in the third formant. The feature 'duration' is split up in long (l) and short vowels (s). The diphthongs are marked with a capital 'D' as they were kept out of the classification. The diphthongs are seen as a separate group because a diphthong exists of two vowels of which the first vowel shades off into the second one.

Table 5-5 *Phonetic transcription of the phonemes (IPA, 1972) with Dutch and British English examples*

Consonant	Dutch	English	Vowel	Dutch	English
/p/	'put'	'post'	/ɑ/	'bad'	'ask'
/t/	'tak'	'table'	/ɛ/	'pet'	'pet'
/k/	'kip'	'kite'	/ɪ/	'fit'	'it'
/f/	'fit'	'fun'	/ɔ/	'pop'	'stop'
/s/	'sap'	'soft'	/i/	'riet	'beat'
/χ/	'giet'	'loch'[1)]	/u/	'koek'	'book'
/b/	'bed'	'ball'	/y/	'duur'	[2)]
/d/	'duur'	'duck'	/œ/	'bus'	'ago'
/v/	'veel'	'very'	/a/	'maan'	[2)]
/z/	'zaai'	'zero'	/e/	'week'	'break'
/m/	'mijn'	'moon'	/o/	'boot'	'coat'
/n/	'naam'	'name'	/φ/	'deuk'	[2)]
/ŋ/	'zang'	'thing'	/ɑu/	'goud'	'cow'
/l/	'leuk'	'long'	/ɛi/	'tijd'	[2)]
/j/	'jas'	'yes'	/ʌy/	'huis'	[2)]
/h/	'hoek'	'have'			
/w/	'weer'	'wine'			
/r/	'rook'	bright[1)]			

[1)] Scottish pronunciation [2)] No English example available

Table 5-6a/b/c*Featrue matrices of the initial consonant, final consonant and vowel*

C1	p	b	t	d	k	f	v	s	z	χ	m	n	l	j	w	h	r
plo	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
fri	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-
nas	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
voi	-	+	-	+	-	-	+	-	+	-	+	+	+	+	+	-	+
pla	b	b	a	a	v	l	l	a	a	v	b	a	a	p	l	g	u

C2	p	t	k	f	s	χ	m	n	ŋ	l	r
plo	+	+	+	-	-	-	-	-	-	-	-
fri	-	-	-	+	+	+	-	-	-	-	-
nas	-	-	-	-	-	-	+	+	+	-	-
voi	-	-	-	-	-	-	+	+	+	+	+
pla	b	a	v	l	a	v	b	a	v	a	u

V	ɑ	ɛ	ɪ	ɔ	i	u	y	œ	a	e	o	φ	ʌy	ɑu	ɛi
F1	1	2	3	2	4	4	4	3	1	3	3	3	D	D	D
F2	b	f	f	b	f	b	m	m	m	f	b	m	D	D	D
F3	u	s	s	r	s	r	r	r	u	s	r	r	D	D	D
dip	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
dur	s	s	s	s	s	s	s	s	l	l	l	l	D	D	D

6 Evaluation of phoneme compression schemes designed to compensate for temporal and spectral masking in background noise

Accepted for publication in Int. J. Audiol.
Goedegebure A, Goedegebure-Hulshof M, Dreschler WA, Verschuure J

Abstract

The effect of phoneme compression on speech intelligibility in background noise has been studied in hearing-impaired listeners with moderate-to-severe high-frequency losses. One configuration (anti-USOM) focuses on a release from spectral masking of high-frequency speech cues by selective spectral tilting. Release from temporal masking is the main goal of a second configuration (HFC), which reduces the speech modulations within a high-pass filtered compression channel. Speech intelligibility was measured with consonant-vowel-consonant (CVC) words in a multi-talker babble and a single-talker background noise. Anti-USOM has a significant negative effect on the phoneme scores in background noise. HFC compression tends to improve vowel intelligibility in a single-talker background noise, especially for the listeners with a relatively poor speech score. In a multi-talker babble noise the effects of HFC compression tend to be negative. It can be concluded that no significant release from spectral or temporal masking is obtained by the applied processing.

6.1 Introduction

Hearing-aid users with a typical high-frequency sensorineural hearing loss often have problems with speech intelligibility in background noise. Restoring audibility by amplifying the high-frequency components usually only partially compensates for the poor identification of speech cues. It is thought that poor cochlear processing at these high frequencies does not allow the perception of all high-frequency speech information. Essential speech cues are inaccessible especially in the presence of a background noise.

A release from *spectral* masking of high-frequency speech cues may be obtained by providing additional tilt in the gain characteristic at high input levels. Positive effects of techniques based on this concept were found for low-frequency band-limited background noises (van

Dijkhuizen et al 1989, Fabry et al 1993, Rankovic et al 1992). The effects can be explained by both increased audibility of high-frequency speech cues and compensation of upward-spread-of-masking (USOM) (Fabry et al, 1993). In speech-shaped noise, however, no convincing evidence has been found that compensation for USOM provides benefits as long as audibility of the high-frequency speech cues is guaranteed. A study of Cook et al (1997) showed no significant improvement of speech recognition in speech-shaped noise after low-frequency gain reduction in hearing impaired listeners with sloping mild-to-moderate losses. Van Buuren et al (1995) indicated that USOM may affect speech intelligibility in wide-band background noise only at high levels and for almost flat frequency responses. Goedegebure et al (2001) showed that adaptive anti-USOM filtering at low frequencies combined with phoneme compression of the high frequencies has a positive effect on vowel perception in conditions without background noise. As the anti-USOM filtering did not provide additional high-frequency gain a release from spectral masking seemed mainly responsible for the effect. Unfortunately, in the same study a remarkable negative effect of anti-USOM filtering was found on both consonant and vowel identification in a multi-talker background noise.

Fast-acting compression is a hearing-aid technique that reduces *temporal* level differences. Theoretically this should result in a release from *temporal* masking of the weaker speech sounds. A possible adverse effect on speech intelligibility may be introduced as well because of a reduced amount of information carried by the speech modulations (Plomp 1988). However, others reject this assumption as *the quality* of the speech information is not necessarily affected by reducing the modulation depth (Villchur 1988). Reducing the modulation depth may even help to normalize the modulation percept as it compensates for the abnormal loudness growth found in perceptive losses.
Although many studies fail to show benefits with compression in background noise (Dillon, 1996), also some positive results have been reported. Moore and Glasberg (1988) found positive effects with a high-frequency compression system and Yund and Buckles (1995) showed improved speech scores with an eight-channel compression system, both at low signal-to-noise ratios and with a stationary speech-shaped background noise. Moore et al. 1999 showed small improvement with fast-acting compression in background noise with spectral or temporal gaps, independent of the number of compression channels. Verschuure et al (1998) found improvements with fast-acting compression in one high-frequency channel for background noises with a fluctuating envelope. However, no positive effect was found in a multi-talker background noise with a similar system (Goedegebure et al, 2001).

The main goal of the present study is to improve the perception of high-frequency speech cues in background noise for hearing-impaired listeners with a sensorineural high-frequency loss. Two signal-processing concepts have been investigated in two separate experiments.

- The first concept focuses on compensation of *spectral-masking* effects by combining phoneme compression with anti-USOM filtering. It is based on the system tested in previous experiments (Goedegebure et al, 2002; Goedegebure et al, 2001) that performed well in conditions without background noise. In the present study we apply the anti-

USOM filtering more moderately as it was suggested that the poor performance in background noise was caused by too much de-emphasis of low-frequency speech cues.

- The second concept is based on the reduction of *temporal-masking* effects. Instead of wide-band control of the compression in the high-frequency channel or a slight high-frequency emphasis of the compression-control signal as tested by (Goedegebure et al, 2001), the compression is now completely controlled by only the high-frequency content of the speech signal. The use of high-frequency control results in a more effective reduction of the speech modulations within the high frequency band. This might be favourable for the perception of weak high-frequency components because of a release of from temporal masking, especially in background-noise conditions with a fluctuating envelope. Only a moderate amount of compression is used (CR=2) to avoid a loss of speech information by too much smoothing of the relevant speech modulations.

Both concepts have been evaluated using speech-intelligibility tests in both stationary and fluctuating background-noise conditions.

6.2 Methods

The speech-processing system

Figures 6-1a, -b and -c show the three compression configurations tested in the current study. Figure 6-1a contains a configuration that has also been used in previous experiments. Essentially it is a two-channel system. In one channel, the so-called compression channel, the high-frequency components of the speech signal are processed by a phoneme compressor. A second channel is used to add the low-frequency components to the compressed high-frequency signal. As no compression is performed in this channel it will be indicated as “the linear channel”. The compression channel contains a high-pass filter to adjust the frequency characteristic according to the pure-tone audiogram of the individual listener. After filtering, the signal is multiplied by the gain calculated in the control channel. The control channel, indicated by the dashed line in figure 6-1a, contains an envelope detector followed by an input-output table. A delay was implemented in the compression channel to compensate for temporal overshoots. The low frequencies are provided by adding the broad-band signal, attenuated by a factor *A*, to the compressed high-frequency signal. An appropriate value of *A* has been determined for each listener to obtain a half-gain frequency characteristic under linear processing. The value of *A* also determined the cross-over frequency between linear and non-linear processing. The choice for an average half-gain characteristic implies that the cross-over frequency was about half-way the slope of the pure-tone audiogram, which was between 1 and 1.5 kHz for most individuals tested in the present study.

A second configuration (figure 6-1b) focuses on a release from *spectral* masking of high-frequency cues. An additional +6 dB/octave high-pass filter was implemented in the linear channel. For low to average input levels the frequency response is quite steep (between mirrored audiogram and half-gain) and no USOM compensation is needed. However, for the higher input levels the frequency response becomes flatter than half-gain and the risk of USOM increases.

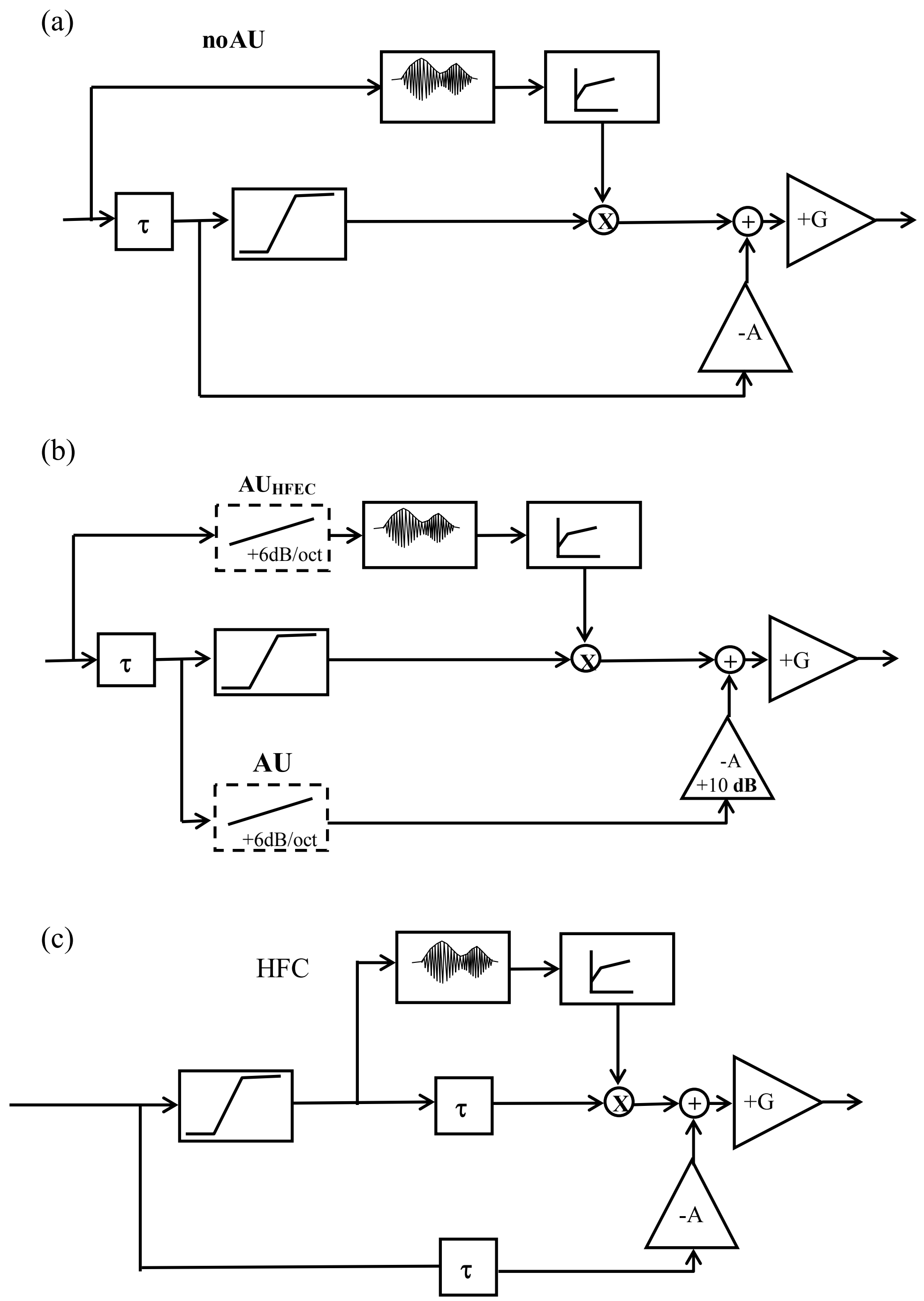


Figure 6-1a/b/c *Simplified schemes of the three compression configurations tested in the current study. All configurations consist of three channels: the compression channel (in the middle), the compression control channel (on top) and the linear channel (below). A delay (τ) is included in all schemes to reduce overshoots caused by the fast compressor. The reference condition is shown in figure (a). The compression signal is filtered by a high-pass filter resembling the mirrored audiogram. The compressor is controlled by the wide-band signal. An attenuated version of the wide-band signal is added to the compressed high-frequency signal. Figure (b) shows the anti-USOM configuration in which the low-frequency content is reduced by a sloping filter (AU) in the linear channel. A value of +10 dB is added to the attenuation value A to compensate for the loss of energy caused by the anti-USOM filter. A second variant is obtained by implementing an additional sloping filter in the control channel (AU_{HFEC}). Figure (c) shows the High-Frequency-Control configuration (HFC) in which the control channel contains the high-pass filtered signal instead of the wide-band signal.*

At these high input levels the additional +6 dB/octave filter dominates the gain at low- and middle frequencies to compensate for possible negative USOM effects. Compared to the system of Goedegebure et al (2001) the attenuation factor *A* in the linear channel was decreased by 10 dB to compensate for the loss of energy caused by the filtering. Figure 6-2 illustrates the effect of the 10-dB compensation on the frequency shaping. The solid line represents a typical example of a frequency response at high input levels, which is flat over a quite large range. Additional filtering to compensate for USOM effects considerably reduces the low- and middle-frequency energy (dashed line), which may result in a loss of available speech information. The 10-dB compensation applied in the linear channel resulted in a more moderate de-emphasis of low-frequency components (dotted line).

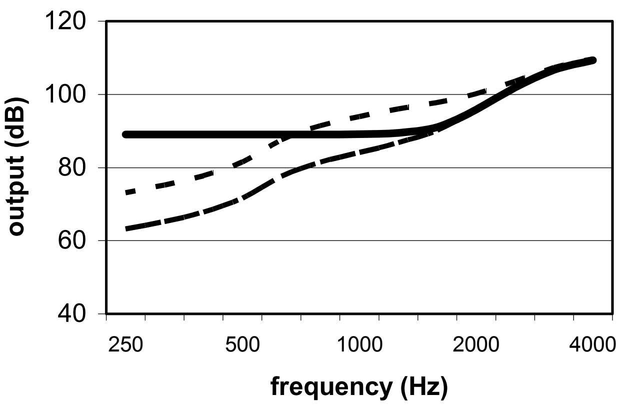


Figure 6-2 *The effect of anti-USOM on the frequency shaping at high input levels. The continuous line shows a typical example of a frequency response at high input levels. Switching on the anti-USOM filter in the linear channel causes a considerable reduction of low- and middle2 -frequency energy in addition (dashed line). A more moderate compensation for USOM is obtained by using a 10-dB level compensation in the linear channel (dotted line)*

A variant of the anti-USOM configuration was obtained by switching on an additional + 6 dB/octave filtering in the control channel of the system (see block HFEC in figure 6-1b). The filter reduces the influence of low frequencies in the compression control and thus counteracts a de-emphasis of relevant vowel cues in the compression channel. This might result in a release from spectral masking of the higher vowel formants by the low formants. The effect of this filter in the control channel was tested in the study of Goedegebure et al (2001) and was shown to improve vowel perception compared to wide-band controlled compression. In the present study it was applied in addition to the anti-USOM filtering, resulting in a configuration that possibly combines the advantages of both approaches.

Figure 6-1c shows a third configuration, called “High Frequency Control” (HFC). The aim of this configuration was to compensate for *temporal* masking of low-level high-frequency cues. The compression-control signal was obtained from the compression channel *after* high-pass filtering, which means that the high-frequency signal was compressed using exactly the same high-frequency signal to control the compressor. This resulted in an effective reduction of modulations within the compression channel (Verschuure et al, 1996). Temporal masking of low-level speech components by preceding high-level components might be reduced as a consequence.
A second possible advantage of reducing the modulations is that in general more high-frequency gain can be supplied without reaching an uncomfortable level. A configuration with extra high-frequency emphasis was therefore included in the tests. In the main HFC-

variant the input-output tables were recalibrated for each listener to obtain a half-gain frequency characteristic at the rms-input level of speech. Extra high-frequency gain was realised in a second variant (HFC+) by *not* compensating the input-output tables for the loss of energy in the control channel caused by the high-pass filtering. An average reduction of 10 dB (the flattest hearing loss) to 20 dB (the steepest hearing loss) was obtained in the control channel, resulting in a 5- to 10-dB extra amplification by the compressor (CR=2).

Experimental set-up

Two experiments were designed to evaluate the benefit with the new speech-processing concepts. In the *first* experiment the following configurations were evaluated:

- conditions REF and noAU2: the configuration without anti-USOM (figure 6-1a) with CR=1 and CR=2 respectively
- conditions AU1 and AU2: the anti-USOM configuration (figure 6-1b) with the +6 dB/octave filtering only in the linear channel and CR set to values of 1 and 2 respectively.
- condition $AU2_{HFEC}$: the anti-USOM configuration (figure 6-1b) with the additional +6 dB/octave filtering in the control channel switched on and CR=2. CR=1 results in the same condition as AU1 and has therefore not been included as an extra condition.

Compression ratios higher than 2 were not tested because poor results were obtained with CR=4 in previous experiments (Goedegebure et al, 2001; Verschuure et al, 1998). A full dynamic-range compression of the speech signal was obtained by choosing a compression threshold of 25 dB below the root-mean-square level of the speech (rms). We used an attack time of 5 ms and a release time of 15 ms. The overshoot delay τ was set to 3 ms.

In the *second* experiment we evaluated the configuration with high-frequency controlled compression (figure 6-1c). The following conditions were tested:

- condition REF: the HFC-configuration with CR=1 and a half-gain frequency response. Note that this condition is exactly the same as the reference condition in the previous experiment.
- condition HFC2: the HFC-configuration with fast compression of CR=2 and time constants of 5 ms (attack) and 15 ms (release). The general amplification by the compressor was chosen such that for each individual the overall frequency response at rms-level of speech was equal to that of the linear reference.
- condition HFC2+: the HFC-configuration with additional emphasis of high frequencies compared to the linear reference. This condition was tested with CR=2, an attack time of 5 ms and a fast release time of 15 ms.
- condition $HFC2+_{slow}$: the HFC-configuration with additional emphasis of high frequencies and a relatively slow release time of 200 ms. This condition was tested with CR=2 and an attack time of 5 ms. It has been included to separate long-term spectral effects from short-term effects as obtained with the fast release time of 15 ms.

Test material

Consonant-vowel-consonant (CVC) word lists were used as speech material in both experiments (Steeneken et al, 1990). Each list consists of 51 CVC words (logatoms) embedded in five different carrier phrases. All consonant-vowel combinations are according to Dutch usage, as are all vowel-consonant combinations. The test words are random combinations of CVC-words, resulting in both nonsense and meaningful words. In total 24 different word lists were used spoken by the same male speaker. The carrier phrase was presented on a screen with an open space where the CVC-word fitted in. The listeners had to repeat the word literally, even if it did not make sense, or if only parts of the word were heard. The repeated words were typed into the computer by the experimenter. After acknowledgement by the listener the typed words were automatically stored.

Two background noises were used with a clearly different temporal behaviour. A multi-talker babble was selected with a large stationary component and small fluctuations in addition (the same track as used by Goedegebure et al. (2001)). Next, a single-talker background-noise was reconstructed from four subsequent sentences of the Dutch speech-in-noise-test spoken by a male speaker (Plomp and Mimpen, 1979). The four sentences were recorded with a Soundblaster 16 Creative Lab Sound Card and the pauses between the sentences were removed. This new signal was looped in the reversed direction and then recorded on DAT. The obtained background noise resembles a continuous discourse of a person speaking an unknown language.

Equipment

The configurations were implemented on a DSP-56001 chip, mounted on an Ariel-board in a PC-configuration. The speech signals and the background noise signals were played by means of two DAT-recorders, controlled by the PC. The level of the background noise was adjusted by a programmable attenuator to obtain the desired SNR-value. The signals were added on an analogue board. After performing anti-aliasing filtering (9 kHz cut-off), the signal were sent to one of the input channels of the Ariel DSP56001-board. The digitized signal was processed by the DSP-processor (sample frequency 22.050 kHz) and converted into an analog signal again. An audiometer (Madsen OB822) was used to adjust the output gain and to present the processed speech signal to the listeners by means of TDH39 headphones with MX/AR41 cushions. The test was performed in a sound proof cabin in which both experimenter and listener were seated.

Listeners

The listeners were selected from the population visiting our Audiological Center by using the following selection criteria:

- a cochlear hearing loss (air-bone gap < 10 dB)
- a maximum speech score < 95% in the routine clinical speech audiogram
- a sloping pure-tone audiogram
- age < 70
- the other ear not more than 30 dB 'better' to avoid cross-talk.

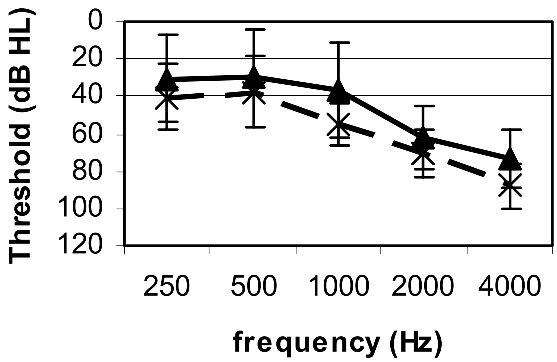


Figure 6-3 *Average pure-tone audiograms of the listeners participating in experiment 1 (solid lines) and in experiment 2 (dotted lines) at the ear that was stimulated. The error bars represent ± 1 standard deviation.*

The tests were performed at our research department. The listeners were paid only for their travel expenses. All tests were completed by 13 listeners in the first experiment and by 11 different listeners in the second experiment. The average pure-tone audiograms of the listeners are plotted in figure 6-3. The solid lines represent the audiograms of the listeners participating in the first experiment, whereas the dotted lines represent the listeners that participated in the second experiment. No explanation is found for the difference in hearing loss between both experiments.

Test procedure

Each of the two aforementioned experiments was performed using three separate test sessions. The contents of the three sessions were similar for both experiments. The first session was mainly used for training purpose and to perform measurements in quiet. The second and third sessions contained the actual speech tests in background noise (test and re-test). Each session lasted about 1 to 2 hours and was performed a week apart for each listener.
In the first session the Most Comfortable Level (MCL) was obtained for the configuration with linear processing. A 13-point loudness scaling method was used with one sentence of the word lists as stimulus (Goedegebure et al, 2001). The gain settings corresponding to MCL were used in all further measurements and all compressor implementations for that particular listener. In experiment 2 a second MCL was obtained for the HFC+ condition and used in further tests with the HFC+ conditions. Then four half wordlists were presented to get the listeners acquainted to the test and to different settings of the system. After that a full CVC-list was presented in quiet for each of the test conditions. The CVC-lists were presented in a fixed order whereas the order of the speech-processing conditions was randomized according to Latin-square schemes. Finally the system was set to the reference condition to measure a signal-to-noise ratio (SNR) for each of the two background noises. An adaptive tracking method was used to obtain a score near a target of 50% phonemes correct (Goedegebure et al, 2001).
In the second and third sessions the CVC-lists were presented at the SNR-levels obtained in the first session. First the phoneme scores in multi-talker babble background noise were obtained for each of the compressor conditions. After a break of at least 5 minutes, the phoneme scores in single-talker background noise were measured. Before the actual measurements took place a word list was presented to acquaint the listener with the task. The

order of the conditions was randomized according to a Latin-squares scheme. Different randomization schemes and CVC-lists were used for the measurements with each of the two background noises. In the third session the procedure of the second session was repeated using the same CVC-lists but different randomization schemes.

6.3 Results

Experiment 1

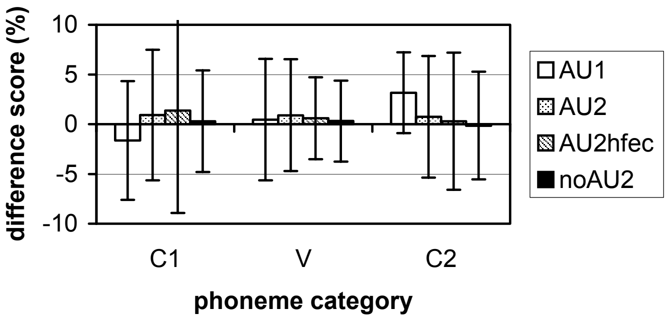


Figure 6-4 *Effects of various processing algorithms on the phoneme score* ***in quiet*** *compared to linear processing averaged across subjects. The results are shown separately for the initial consonant (C1), the vowel (V) and the final consonant (C2) respectively. The error bars represent ± 1 standard deviation.*

In the first experiment effects of combined phoneme compression and anti-USOM were evaluated. Figure 6-4 shows the effects of processing on the phoneme score in quiet compared to the reference condition. Only marginal differences are found, except for an increased final-consonant score with anti-USOM. However, no such increase is found in the conditions with combined anti-USOM and compression.

Table 6-1 *Statistically significant effects in experiment 1,* p<0.05, ** p<0.01*

condition	*main effects & interactions*	*F-value*	*contrasts*
quiet	phoneme category**	F(2,12)=16.6	V> C2,C1 **
noise	phoneme category **	F(2,12)= 132.2	V>C2,C1** C2>C1*
	anti-USOM *	F(2,12)= 3.4	AU, AU_{hfec} < Ref *

To analyze the effects of signal processing we performed a repeated-measures ANOVA with within-subject factors *phoneme category* (initial consonant, final consonant or vowel), *compression* (CR=1 or CR=2) and *anti-USOM* (noAU, AU or AU_{HFEC}). Significant effects are listed in table 6-1. No significant effects have been found for *compression* and *anti-USOM* (p>0.05), indicating that both signal-processing concepts did not substantially change the performance in quiet. A significant effect has only been found for the factor *phoneme category* (F(2,12)=16.6, p<0.05). This is caused by a higher vowel score (88.4% on average) compared to the initial-consonant and final-consonant score (both 81.5% on average).

The effects of speech processing in background noise are presented in figures 6-5a and 6-5b. The figures show the difference in phoneme scores compared to linear processing in multi-talker babble and single-talker noise respectively. The *absolute* phoneme-scores are between

50 and 60% on average in all background noise conditions, which is just above the target score of 50% phonemes correct for each individual. In addition, table 6-1 lists the significant main effects and interactions as obtained by a repeated-measures ANOVA with within-subject factors *background-noise* (multi-talker babble or single-talker), *time* (test or re-test), *phoneme category, compression* and *anti-USOM*.

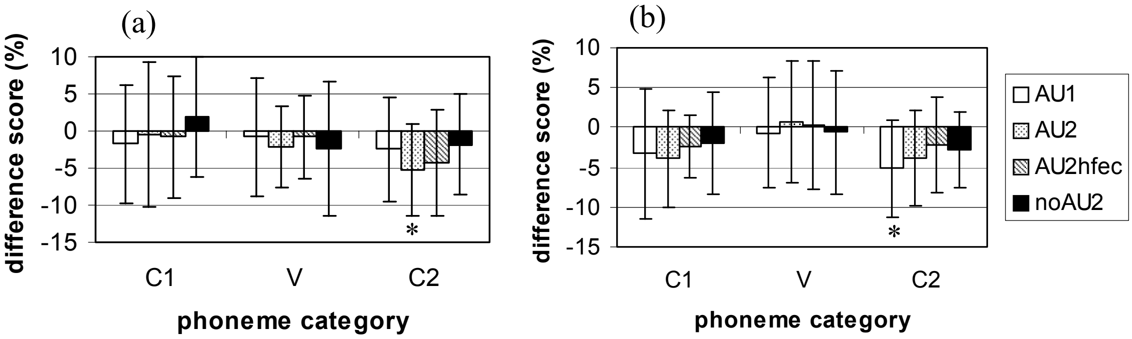


Figure 6-5a/b *Effects of various processing algorithms on the phoneme score in multi-talker babble (a) and single-talker babble (b) compared to linear processing averaged across subjects. The results are shown separately for the initial consonant (C1), the vowel (V) and the final consonant (C2) respectively. The error bars represent ± 1 standard deviation. Significant effects are marked (* paired t-test, p<0,05)*

Remarkable is the significant negative effect of anti-USOM ($F(2,12)= 3.4$, $p<0.05$) as found by the repeated-measures analysis. Figures 6-5a and 6-5b show that the negative effect of anti-USOM is mainly explained by a poorer performance on consonant perception. In multi-talker babble noise the largest differences have been found for the final-consonant, with a significant negative effect for the combination of compression and anti-USOM (paired t-test, $p<0.05$ with Bonferroni-correction for multiple tests). In the single-talker noise the performance with anti-USOM tends to be poorer for both initial- and final consonant intelligibility. The only significant effect we have found is the negative effect of anti-USOM filtering on the final-consonant score (paired t-test, $p<0.05$ with Bonferroni-correction for multiple tests).

The repeated-measures analysis did not reveal a significant effect of compression ($F(1,12)=1.2$, $p>0.05$). Figure 6-5a shows that the initial-consonant score in multi-talker babble noise tends to improve with phoneme compression only (noAU2). However, the effect is counteracted by a decrease of the final-consonant and vowel score. It is disappointing that the performance in single-talker noise tends to decrease as well by the use of phoneme compression.

Table 6-1 shows that *phoneme category* is also a significant factor ($F(2,12)=132.2$, $p<0.001$). In both background noises the vowel score exceeds the final-consonant score by about 14 %, whereas the final-consonant score exceeds the initial-consonant score by 11 to 12%. The masking effect of the background noise seems to be stronger for initial-consonant perception than for vowel- and final consonant perception.

To identify consistent differences between subjects we repeated the analysis as described in the previous paragraph with a number of additional between-subject factors: *PTA* (pure-tone average 500 Hz, 1kHz, 2 kHz and 4 kHz), *slope* (difference between thresholds at 4 kHz and

500 Hz), CVC_{quiet} (phoneme score in quiet under linear processing), SNR_{mt} (signal-to-noise ratio in multi-talker noise, obtained from adaptive tracking procedure in the first session), and SNR_{st} (signal-to-noise ratio in single-talker noise, obtained from adaptive tracking procedure in the first session). No significant interactions have been found between the within-subject factors *anti-USOM* and *compression* and any of the defined between-subject factors. This means that the effects of signal processing are not related to specific between-subject characteristics.

Experiment 2

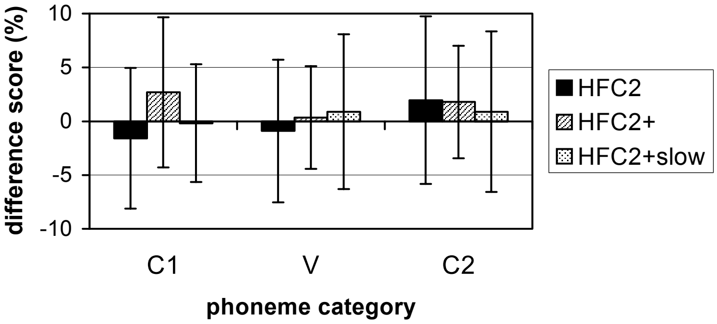


Figure 6-6 *Effects of processing on the phoneme score in quiet compared to linear processing averaged across subjects. The results are shown separately for the initial-consonant (C1), vowel (V) and final-consonant (C2) respectively. The error bars represent ± 1 standard deviation.*

In the second experiment the performance with high-frequency controlled (HFC) compression was evaluated. The effect of HFC-processing in quiet is presented in figure 6-6. In addition a repeated-measures analysis was performed with factors *phoneme category* (initial consonant, final consonant or vowel) and *compression* (each of the three HFC conditions and the linear condition). No significant effects of processing have been found. Like in the previous experiment, a significant effect has been found for the factor *phoneme category* ($F(2,10)=20.1$, $p<0.001$) as the vowel score is higher than the initial- and final-consonant score (84% against 76% and 74% respectively).

Table 6-2 *Statistically significant effects in experiment 2, * p<0.05, ** p<0.01, n.a. = not applicable*

condition	*main effects & interactions*	*F-value*	*contrasts*
quiet	phoneme category**	F(2,10) = 20.1	V > C2,C1**
noise	phoneme category**	F(2,10) = 122.1	V>C2,C1** C2>C1*
	time*	F(1,10) = 6.0	re-test > test*
	HFC x phoneme category *	F(6,10) = 2.8	n.a.
	noise x phoneme categroy*	F(2,10) = 6.6	n.a.
	HFC x noise x phoneme category*	F(6,10) = 2.3	n.a.

Figures 6-7a and 6-7b show the effects of HFC-processing in multi-talker babble and single-talker background noises respectively. The bars represent the difference between the processed condition and the linear reference condition for initial-consonant, vowel- and final-consonant respectively averaged across listeners. As in the previous experiment, the *absolute* phoneme scores are between 50 and 60% on average in all background noise conditions. In addition, table 6-2 lists the main significant within-subject effects and interactions as obtained by a repeated-measures ANOVA with within-subject factors *compression, phoneme category, time* (test or re-test) and *noise* (multi-talker babble or single-talker background noise).

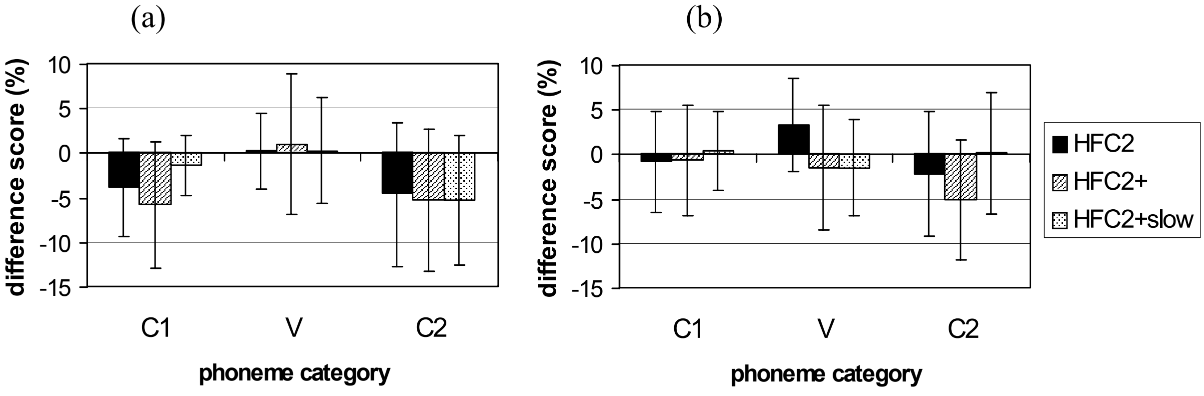


Figure 6-7a/b *Effects of processing on the phoneme score in multi-talker babble (a) and single-talker babble (b) compared to linear processing averaged across subjects. The results are shown separately for the initial-consonant (C1), vowel (V) and final-consonant (C2) respectively. The error bars represent ± 1 standard deviation.*

Although no significant main effect of compression has been found (see table 6-2), figures 6-7a and 6-7b show some remarkable differences compared to linear processing. In the single-talker noise (figure 6-7b) the use of HFC tends to improve vowel perception compared to linear processing. Unfortunately no such positive tendency is found for the consonant score. In multi-talker babble noise the initial- and final-consonant scores decreases with compression, while no difference is found for the vowel score. Apparently, HFC processing has a different effect on the vowel score than on the consonant score. This is confirmed by the significant two-way interaction found between *compression* and *phoneme category* ($F(6,10)=2.8$, $p<0.05$). The effect depends also on the noise condition as a significant three-way interaction has been found between *compression*, *phoneme category* and *noise* ($F(6,10)=2.3$, $p<0.05$). No large differences are found between the three different HFC-conditions. Using a slower release time (condition HFC2+$_{slow}$) results in a performance that is very similar to that measured with linear processing, except for a decreased final-consonant score in multi-talker babble noise. Furthermore, the performance with HFC-compression does not improve with additional spectral tilting (HFC+ compared to HFC). In single-talker noise the score even tends to decrease as a result of the additional tilting. In general the additional tilting applied in the HFC+-conditions hardly seems to influence the performance with HFC compression.

Next to the effects of signal processing, significant main effects have been found for factors *phoneme category* ($F(2,10)=122.1$, $p<0.001$) and *time* ($F(1,10)= 6.0$, $p<0.05$). In both background noises the vowel score is higher than the final-consonant score (+18.3% on average) while the final-consonant score is higher than the initial-consonant score (+12.5% on average). Interesting is that in single-talker noise the vowel scores are a bit poorer and the consonant scores are slightly better compared to the corresponding phoneme scores in multi-talker noise. This results in a significant interaction of *phoneme category* and *noise*

(F(2,10)=6.6, $p<0.01$). The effect for factor *time* is caused by a better performance in the re-test session (+1.9% on average). Note that no significant interaction has been found between *time* and *HFC*. This means that repeating the measurement did not influence the effect of HFC on phoneme score. Therefore we did not analyze the effects of HFC in test- and re-test separately.

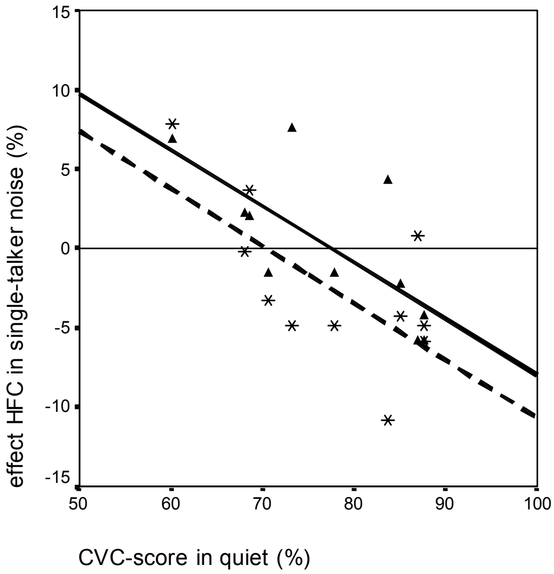


Figure 6-8 *The effect of fast HFC-compression on the phoneme score in single-talker noise as function of the phoneme score in quiet, for both conditions HFC2 (triangles) and HFC2+ (stars). A linear fit is plotted for each condition with correlation coefficients of -0.71 (HFC2, continuous line) and -0.68 (HFC2+, dashed line), both with p<0.05.*

The repeated-measures analysis was repeated with the between-subject factors as defined in the previous experiment (*PTA*, *slope*, CVC_{quiet}, SNR_{mt} and SNR_{st} respectively). The only two-way interaction with within-subject factor *HFC* has been found for CVC_{quiet} (F(3,10)=4.1, $p<0.05$). To investigate the relation between both factors we correlated the overall phoneme score in quiet with the effects of HFC-processing compared to linear processing. In general a negative correlation has been obtained, especially for the two conditions using fast time constants. Figure 6-8 shows this negative correlation with speech discrimination for conditions HFC2 (triangles and continuous line) and HFC2+ (stars and dashed line) in single-talker noise. The corresponding correlation coefficients are -0.71 and -0.68 respectively ($p<0.05$). It means that most benefit from HFC-processing has been found for the listeners with relatively poor speech discrimination.

6.4 Discussion

In the present study we investigated two phoneme-compression concepts to improve the identification of high-frequency speech cues in background noise. One concept ("anti-USOM") uses additional tilting of the low-frequency tail of the spectrum to compensate for spectral-masking effects. A second concept consists of high-frequency controlled compression (HFC compression) to compensate for temporal-masking effects.

Effects of anti-USOM

A significant negative effect of anti-USOM has been found on phoneme intelligibility in background noise (repeated measures ANOVA, $p<0.05$). The main negative effects have been found for consonant perception in a multi-talker babble. Although the effect of anti-USOM is negative, the size of the effect has been reduced compared to the study of Goedegebure et al (2001). The most important difference is that the low-frequency de-emphasis has been applied more moderately in the present study. This supports our initial assumption that the poor performance with anti-USOM in background noise is caused by de-emphasis of relevant low-frequency cues.

Unfortunately the resulting effect of anti-USOM remains negative in background noise. The negative effect caused by de-emphasis of low-frequency cues seems to dominate a possible enhancement of high-frequency cues due to compensation for USOM. It suggests that the residual effect of USOM is only small after the speech-spectrum has been adjusted to the hearing loss according to a half-gain rule. This is in line with results of other studies that show hardly any effect of spectral shaping in conditions with stationary speech-shaped background noise as long as audibility has been taken in account (Van Buuren and Festen, 1994; Cook et al, 1997). Therefore, the possible benefits that can be obtained by signal-processing techniques based on balance adjustments seem to be limited in speech-shaped background noise.

Another interesting difference compared to the experiment of Goedegebure et al (2001) is that we have found hardly any positive effect of anti-USOM on vowel perception in quiet. The moderate low-frequency de-emphasis is probably not enough to compensate for masking effects of higher formants by the first formant. Summers and Leek (1997) also showed that a considerable attenuation of the first formant is needed to improve vowel perception in quiet. These results suggest the use of a kind of adaptive low-frequency suppression that anticipates the acoustical condition. In relatively quiet conditions a high amount of suppression should be provided while in background noise the low-frequency suppression should be applied more moderately. Interestingly, most hearing-aid strategies act in an opposite way as low frequencies are suppressed mainly in noisy conditions. This might be a good approach for low-frequency background noises like traffic noise. However, for speech perception in a broad-band background noise the hearing-impaired listeners with poor cochlear high-frequency processing mainly rely on the low-frequency cues (Goedegebure et al. 2002). In that particular case de-emphasis of the low frequencies is likely to affect speech perception instead of improving it. In quiet, on the other hand, a substantial attenuation of low-frequency energy has negligible effect on the perception of low-frequency cues whereas the perception of high-frequency cues may be improved.

Effects of high-frequency controlled compression

The use of fast-acting compression with high-frequency control (HFC) did not reveal evident benefits on speech intelligibility in background noise either. However, there is no significant overall negative effect as found with the anti-USOM concept in the first experiment. In the single-talker background noise vowel intelligibility even tends to improve with HFC-compression. The reduction of temporal envelope-fluctuations within the compression channel seems to improve the identification of vowel cues. The improvements have mainly been found for listeners with relatively poor speech discrimination in quiet, as a negative correlation has been found between the effects of HFC-processing and the phoneme score in quiet. The poorer phoneme score for these listeners may be caused by a poorer temporal resolution, which would explain a better performance with phoneme compression in a fluctuating background noise. This would be in contrast to the results of Turner et al (1995), who found no relationship between amount of hearing loss and impairment of temporal acuity. Interestingly, the HFC processing seems to be mainly effective for vowel perception and not for consonant perception. Possibly the relatively weak high-formant cues can be identified more easily due to the reduced peak-levels of the noise. The compensation of temporal masking realized by the phoneme compressor seems however not strong enough to improve the perception of the more subtle consonant cues.

In multi-talker babble noise the difference in performance between consonant and vowel perception is remarkable. Compression tends to affect both initial- and final consonant intelligibility while the effect on vowel intelligibility is negligible. A possible negative effect of compression may be caused by the attenuation of consonants that contain a relatively high amount of high-frequency energy, like fricatives. The fast release time used for the compression (15 ms) seems mainly responsible for the aforementioned differences in speech intelligibility. For the condition with a slower release time (200 ms) hardly any differences have been found compared to linear processing. Additional tilting of the spectrum did not lead to substantial changes in performance as well. The effects measured with the phoneme compressor in this experiment were therefore mainly determined by temporal aspects.

Influence of temporal characteristics of background noise

Both experiments show that the effect of the applied signal-processing depends on the acoustical conditions. In a more stationary background noise such as multi-talker babble, both anti-USOM and phoneme compression seem to deteriorate phoneme perception instead of improving it. Both types of processing should therefore be applied with care to avoid substantial negative effects in stationary background noise. In a background noise with a fluctuating character, however, the negative effects have been less prominent and even a slight benefit could be obtained using high-frequency controlled compression.

Verschuure et al (1998) also found that the temporal characteristic of the background noise determines the effect found with phoneme compression. One possible reason is that fast-acting compression may compensate for temporal masking caused by the noise. Temporal masking of the speech mainly occurs when the background noise has a fluctuating envelope. Certain speech cues within the ‘gaps’ of the noise cannot be perceived because they are masked by the preceding noise ‘burst’. Therefore, the higher the amount of fluctuations, the higher the chance that fast-acting compression compensates for temporal masking. A study of Moore et al (2001) with gap detection in noise confirms that temporal-masking effects caused by modulations up to 50 Hz can indeed be reduced by fast-acting compression. This would require a multi-channel compression algorithm that effectively reduces the temporal fluctuations within the various frequency bands. Several studies that show positive effects of fast-acting compression in background noise indeed use some kind of multi-channel technique (Yund and Buckles (1994), Moore et al (1999)). In our study the positive tendency found for improved vowel perception in fluctuating background noise also suggests the use of a compression algorithm that reduces intensity differences within the high-frequency band.

6.5 Conclusions

- listeners with perceptive high-frequency sloping losses do not profit from anti-USOM processing as applied in our compression system, even when applied moderately. Instead of the expected release of masking a significant negative effect has been found on speech intelligibility in background noise.
- No significant compensation for temporal-masking effects has been obtained by compression of speech and noise modulations within a high-pass filtered compression channel. However, a tendency has been found for improved vowel perception in a single-talker background noise.
- the temporal character of the background noise partly determines the effect of phoneme compression on speech intelligibility.

6.6 Acknowledgements

This research has been performed in the framework of the HEARDIP-project and the SPACE-project, both funded by the European Commission. We want to thank our European colleagues for their contribution to this study by the fruitful discussions and cooperation within the consortium. We also acknowledge the contributions made by the Heinsius Houbolt Foundation to support this research.

6.7 References

van Buuren RA, Festen JM, Plomp R (1995) Evaluation of a wide range of amplitude-frequency responses for the hearing impaired. J Speech Hear Res 38: 211-221.

Cook JA, Bacon SP, Sammeth CA. (1997) Effect of low-frequency gain reduction on speech recognition and its relation to upward spread of masking. J Speech Lang Hear Res 40:410-422.

van Dijkhuizen JN, Festen JM, Plomp R (1989) The effect of varying the amplitude-frequency response on the speech-reception threshold of sentences for hearing-impaired listeners. J Acoust Soc Am 86(2): 621-628.

Dillon H. (1996) Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response time. Ear Hear 17:287-307.

Fabry DA, Leek MR, Walden BE, Cord M (1993) Do adaptive frequency response (AFR) hearing aids reduce 'upward spread' of masking? J Rehab Res 30 (3) : 318-325.

Goedegebure A, Hulshof M, Verschuure J, Dreschler WA. (2001) Effects of Single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners. Audiology 40(1):10-25.

Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA. (2002) The effects of phonemic compression and anti-upward-spread-of-masking (anti-USOM) on the perception of articulatory features in hearing-impaired listeners. Int J Audiol 41(7): 414-428.

Moore BCJ, Glasberg BR. (1988) A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. Brit J Audiol 22:93-104.

Moore BC, Peters RW, Stone MA. (1999) Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J. Acoust Soc Am 105 (1): 400-411

Moore BCJ, Glasberg BR, Alcantara JI, Launer S, Kuehnel V. (2001) Effects of slow- and fast-acting compression on the detection of gaps in narrow band noise. Br J Audiol 35:365-374.

Plomp R, Mimpen AM. (1979) Improving the reliability of testing the speech reception threshold for sentences. Audiology 18(1):43-52.

Plomp R (1989) The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function. J Acoust Soc Am 83 (6): 2322-2327.

Rankovic CM, Freyman RL, Zurek PM. (1992) Potential benefits of adaptive frequency-gain characteristics for speech perception in noise. J Acoust Soc Am 91(1):354-362.

Steeneken HJM, Geurtsen FWM, Agterhuis E (1990) Speech data-base for intelligibility and speech quality measurements. TNO-Institute for Perception: Soesterberg.

Summers V, Leek MR. (1997) Intraspeech spread of masking in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 101(5):2866-2876.

Turner CW, Souza PE, Forget LN (1995) Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J Acoust Soc Am 97 (4): 2568-76.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP. (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Verschuure J, Maas AJJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA (1996) Compression and its effect on the speech signal. Ear Hear 17:162-175.

Villchur (1989) Comments on " The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function". J Acoust Soc Am 86 (1): 425-427

Yund EW, Buckles KM. (1995) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239.

7 Phoneme compression in experimental hearing aids: effects of everyday-life use on speech intelligibility.

in revision Int. J. Audiol.
Goedegebure A, Quartel-Droogendijk JM, Dreschler WA,Verschuure J

Abstract

We have evaluated the effect of phoneme compression on speech intelligibility in hearing-aid users during a period of regular use of the system. Phoneme compression was applied in one high-frequency channel and low-frequency information was supplied by a second linear channel (Goedegebure et al, 2005; Goedegebure et al, 2001). Slow-acting non-linear processing was used additionally to keep speech of various levels within the active range of the phoneme compressor. Six experienced hearing-aid users tested the various programs in everyday environments using a digital wearable device. Speech tests were performed every week at the hospital.
The resulting effect of phoneme compression on speech intelligibility tends to be better in quiet than in background noise. Interestingly, speech intelligibility with the compression programs increased considerably after regular use in everyday life. The increase is slightly stronger for combined slow-acting and phoneme compression compared to slow-acting compression only. These results suggest that acclimatisation to phoneme compression may have slightly influenced the performance.

7.1 Introduction

It is almost impossible to imagine present-day hearing-aid systems without non-linear amplification. The possibilities to implement compression techniques in hearing aids have been increased considerably since the introduction of digital processing. Compression is mainly used to improve listening comfort for hearing impaired listeners with a reduced dynamic range. It compensates for the faster growth of loudness (recruitment) and thus avoids that soft sounds become inaudible and loud sounds become uncomfortably loud. This can be achieved in slow-acting compression systems. Some systems, however, use very fast time constants to reduce the short-term envelope fluctuations in speech. The aim of these "phoneme compression" systems is to emphasise the weak speech components in order to optimise speech discrimination.

So far laboratory experiments have not shown convincing evidence for the benefit of phoneme compression. Few studies found improvements with phoneme compression compared to an optimised linear system (Yund and Buckles 1995a, Moore et al. 1999, Verschuure et al. 1998, Goedegebure et al. 2001) but numerous studies showed no or even adverse effects (e.g. van Buuren et al. 1999, de Gennaro et al. 1986, van Harten-de Bruijn et al. 1997, Kollmeier et al. 1993, Marzinzik et al. 1997, Moore et al. 2004, Souza and Bishop 2000, Walker et al. 1984).
In most clinical studies no general preference has been shown for either phoneme compression or slow-acting compression when comparing both types of processing (Appel et al. 2002, Bentler and Nelson 1997, van Toor et al. 2002). However, van Toor et al. 2002 found a good performance with phoneme compression in background noise within the subgroup that preferred this type of processing. Gatehouse et al. (1999, 2003) confirms that certain groups of listeners may profit from phoneme compression and others may not. They found a higher benefit from phoneme compression for listeners with a greater cognitive ability and a more active life style.

Using phoneme compression in clinical practice includes some additional factors as compared to evaluation in a laboratory condition. The system should compensate for loudness recruitment by reducing the overall level differences between various kinds of sounds. Listening comfort becomes also a relevant topic. This may have consequences for the design of the system, such as using a combination of different types of compression as introduced by Moore and Glasberg (1988). Another relevant factor that may influence the results is the time needed by the auditory system to adapt to the processed speech, known as perceptual acclimatisation. It is known that the central auditory system is a functionally and physiologically plastic system (Palmer et al. 1998), although there is some discussion about the amount of plasticity (Turner and Bentler 1998). Deprivation of the central auditory system plays a role as a result of a persistent hearing loss with an inadequate hearing-aid solution (Arlinger et al 1996). Acclimatisation to amplified speech has been found in several studies (Gatehouse 1992, Gatehouse 1993, Cox et al 1996, Munro and Lutman 2003) although others found little evidence for improvements over time (Saunders and Chienkowski 1997, Gabriel 2000, Keidser and Grant 2001, Humes and Wilson 2003, Humes et al. 2004). Some studies show that auditory acclimatisation also may play a role when non-linear processing or speech processing is applied into hearing aids (Arlinger and Billermark 1999, Kuk et al. 2003, Yund and Buckles 1995b). This suggests that it has an additional value to evaluate the benefit from non-linear processing after a period of using a device with this type of processing. Laboratory studies usually do not include such an acclimatisation period before testing.

In the present paper non-linear hearing-aid processing has been tested under laboratory conditions and in everyday environments by a group of hearing-impaired listeners. Phoneme compression is applied within a high-frequency channel whereas the low frequencies are processed linearly (Goedegebure et al 2001, Goedegebure et al 2005). The phoneme compressor is embedded in a slow-acting non-linear system to maintain speech of different levels within the working range of the phoneme compressor and thus adapt the system for use in everyday environments. The main goal of the study is to evaluate the effect of phoneme compression on speech intelligibility during a period of regular use of the system by hearing impaired listeners.

7.2 Design and implementation of the hearing-aid system

Phoneme compression schemes

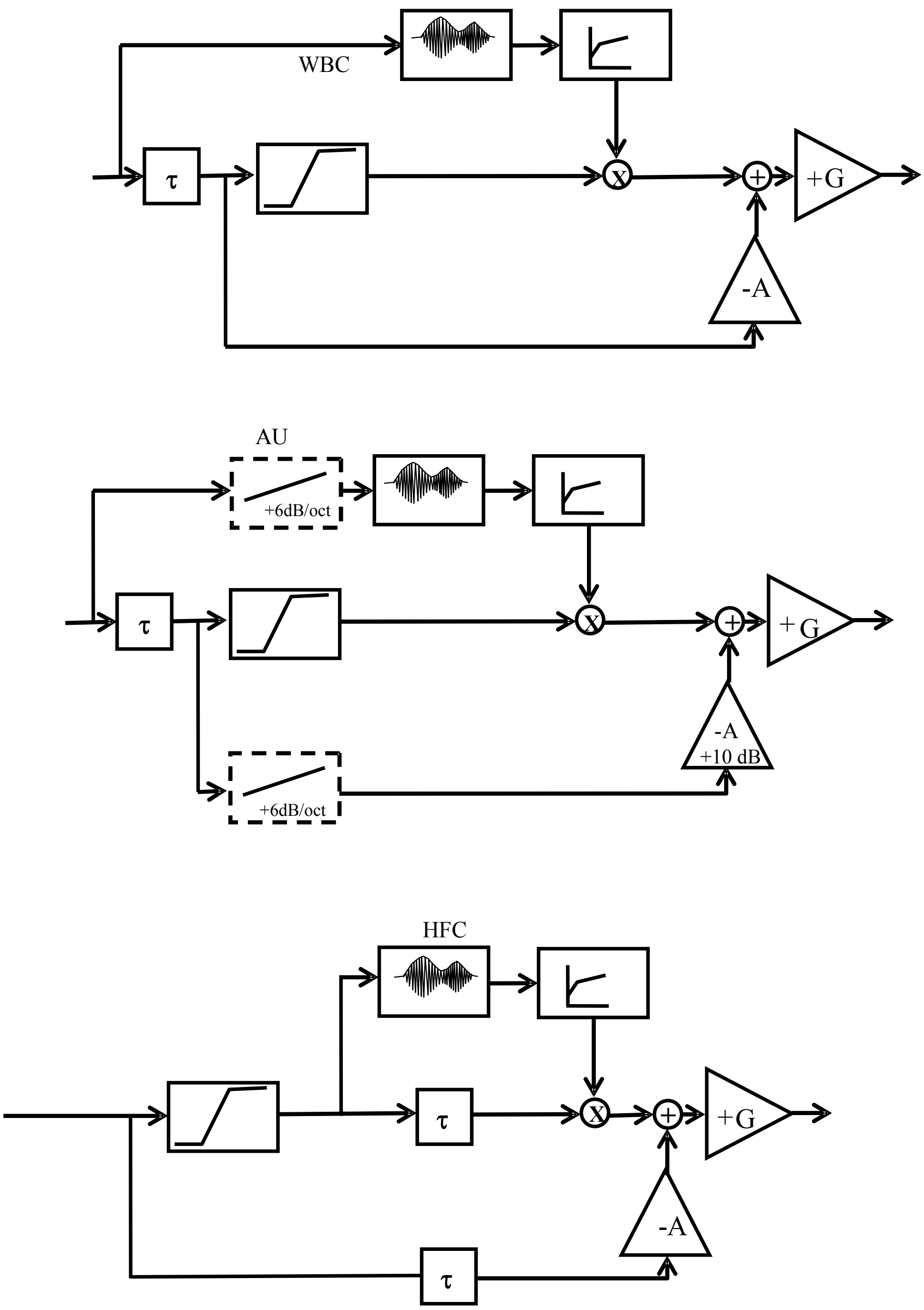


Figure 7-1a/b/c *Simplified block schemes of the three compression configurations tested in the current study.*

The main part of the system consists of a phoneme compressor in a high-frequency channel (Goedegebure et al 2005, Goedegebure et al 2001). In the present study three configurations have been tested. The simplified block schemes of each of the configurations are presented in figures 7-1a to 7-1c. The first configuration is called "Wide-band-Control" (WBC). The speech signal is filtered by a high-pass filter and then compressed by a fast-acting compressor. The filter is adjusted to the hearing loss of the listener by mirroring of the pure-tone audiogram. The level of the full speech signal, the wide-band signal, controls the compression of the high frequencies. The low frequencies are supplied by adding the attenuated wide-band signal to the compressed signal. On average a half-gain frequency characteristic is obtained. The second configuration is called Anti-Upward-spread-of-masking (AU) because two sloping high-pass filters have been added to compensate for upward-spread-of-masking effects. The third configuration is called High-Frequency Control (HFC) as the compressor is now controlled by only the high-pass filtered part of the speech signal.

The differences between the three configurations are in short:

- Configuration WBC emphasises low-level consonant cues and reduces the vowel energy within the high-frequency channel.
- AU additionally compensates for spectral-masking effects by reducing the high-level low-frequency speech components.
- HFC mainly focuses on the reduction of temporal-masking effects as it reduces intensity differences within the high-frequency band.

Adapting phoneme compression to everyday environments

The phoneme compression schemes as described in the previous paragraph needed adjustment for use in everyday environments. Instead of speech at one fixed input level, as used in previous laboratory experiments, speech with a broad range of input levels should be handled. Extending the dynamic range of the phoneme compressor to 70 or 80 dB is undesirable, as low-level environmental and system noises are highly amplified within silent intervals between words ("pumping"). Another possible solution would be to implement a slow-acting compressor (AGC) in front of the phoneme compressor, such as the dual-front end system of Moore and Glasberg (1988). However, a high amount of slow-acting compression is needed to keep speech signals of various input levels fully within the small 45-dB dynamic range of the phoneme compressor. As a consequence differences between soft and loud sounds will be diminished and may not be perceived by the hearing-aid user.

To deal with this problem we added a third non-linear system at the output of the phoneme compressor. Figure 7-2 shows a simplified block diagram of the complete system. The first block consists of an AGC that uses a high amount of compression over a broad dynamic range. The value of the compression threshold is very low (40 dB SPL) and the compression ratio is very high (CR=16). As a result, all speech signals with levels above 40 dB SPL are mapped to a fixed level in the middle of the dynamic range of the phoneme compressor. After that the phoneme compressor performs a full-dynamic range compression of each of the speech signals. A third non-linear block is implemented in order to partly restore the original overall level differences. It consists of an *expander* that is driven by the *same* envelope detector as used for the AGC. This synchronised expander generates a gain value by which the output sample of the fast compressor is multiplied. In this way, the high amount of compression applied at the beginning can be accurately compensated at the end of the system. Any desired input-output characteristic can be obtained by choosing an appropriate shape of the input-output curve of the expansion block (for details, see Appendix 7-10). An essential part of the system is the synchronisation between the envelope values used to calculate the gain in both the AGC and the expansion system. Without this synchronisation the system may become unstable because of the extreme high compression ratio of 16 used in the slow-acting compressor.

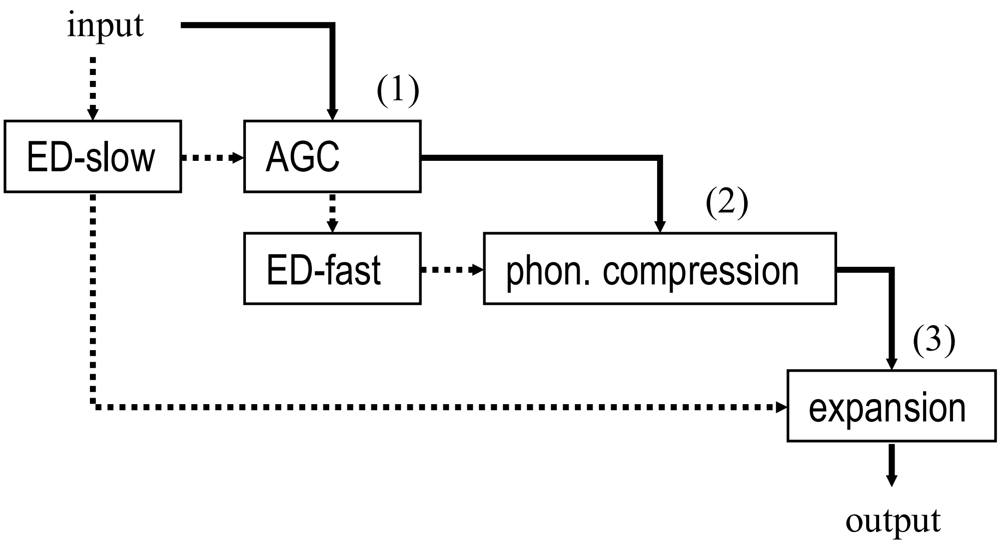


Figure 7-2 *A schematic representation of our compression system with three stages. The dotted arrows represent the control-signal flow. The main signal flow is represented by the continuous lines. The input signal first passes an AGC (1) with a slow-acting envelope detector (ED-slow). After that the phoneme compression takes place (2), using a fast envelope-detection system (ED-fast). As a third step expansion is performed (3), using the envelope signal of the slow-acting AGC. The synchronised expander partly compensates for the reduced dynamic range applied by the slow-acting AGC in front of the system (1).*

Settings of the compression parameters

Figure 7-3 shows the resulting input-output characteristics used in our experiments for the slow-acting non-linear processing blocks. The output level of the resulting characteristic in this figure is arbitrarily set at 30 dB gain for an input level of 60 dB SPL, but it can be set to any value required by the listener.

The shape of the resulting input-output characteristic is divided in three parts:

- between the input levels of 42 SPL and 76 dB SPL slow-acting compression is performed with an effective compression ratio CR_{res}=2 to have smoothed compensation for the recruitment of the impaired ear. The resulting compression is achieved by using CR_1=16 in the front-end stage and CR_3=16/23 in the expansion block (see equation (7.6) of the Appendix 7-10);

- above the input level of 76 dB SPL the CR_l=16 of the slow-acting AGC determines the resulting amount of compression. This means that the system acts as a soft input limiter above input levels of 76 dB SPL;
- below the input level of 42 dB SPL we use expansion with a factor 2.5. We have two important reasons to implement low-level expansion. The first reason is to obtain a reduced sensitivity to acoustical feedback with low gain levels at silent intervals. The second reason is that compression emphasises low-level system noise and soft environmental sounds, which may be annoying and disturbing to the listener. Low-level expansion reduces the amplification of these sounds in the relatively silent intervals.

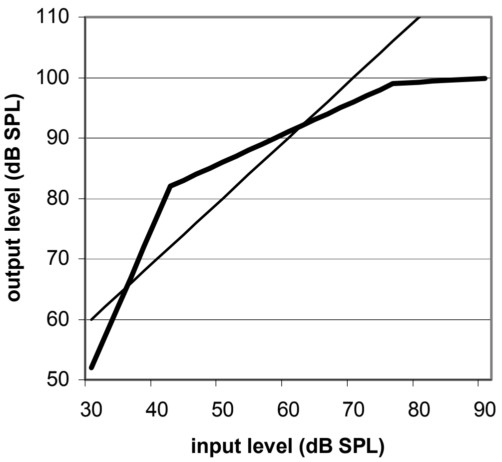


Figure 7-3 *The resulting input-output characteristic of the slow-acting AGC and expander as used in the field study (thick line). The thin line represents a linear input-output relation with a fixed 30 dB gain.*

The AGC in front is based on the dual-front end system from Moore and Glasberg (1988) and the software was delivered by their department. Only the slow compressor was used, which is characterised by three time constants: attack (310 ms), holdoff (500 ms) and release (800 ms). The holdoff circuit holds the AGC gain constant for a short time after the signal has dropped below the level necessary to cause the AGC to reduce the gain further. Then the AGC gain ramps back linearly in the same fashion as before. This avoids pumping between words in a sentence.

The input-output characteristic of the fast compressor was set to an active range of 45 dB and a compression ratio of 2. The time constants were set to 5 ms (attack) and 15 ms (release). A delay of 3 ms was used to reduce temporal overshoots. As reference condition we used configuration WBC with a compression ratio of 1 for the fast compressor. The differences in long-term frequency response between the compression conditions and linear processing have been minimised by calibrating the input-output characteristics with speech-shaped noise. An average half gain characteristic was used as a target. The anti-upward-spread-of-masking condition AU was characterised by a slightly deviating characteristic in all cases because of the sloping filter in the low-frequency channel. A digital offset at the end of the system compensated for the measured differences in overall level between the three different compression schemes.

Implementation of the programs

The programs have been implemented on a digital wearable unit (Rass 1996) with a Motorola 56001 Digital Signal Processor. Two parallel non-linear systems were implemented in each program to achieve independent processing for both left- and right-ear signals. The sampling frequency of the system is 16 kHz. However, for two subjects we had to use 12 kHz because at 16 kHz the steep filters required too much calculation time for one clock cycle. A digital volume control has been implemented by a look-up table at the end of the system, consisting of 10 subsequent steps of 2 dB that could be controlled by the volume switch on the wearable unit. Analogue earpieces (modified Siemens 584-PGC Behind-The-Ear's (BTE's)) receive the signal, transmit it to the digital device and transmit the processed signal back into the ear. We set the balance at the BTE's to 'Normal' and the output to a maximum of 125 dB SPL. The earhook contains Knowles acoustic damping systems (green, 1500 Ohms) to obtain a relatively flat frequency response. The power supply consists of a battery pack that has to be recharged after 8 hours of use.

7.3 Experimental design

Listeners

The selection of listeners was based on the following criteria:

- Symmetrical sloping sensorineural high-frequency losses
- Maximum speech discrimination in the routine clinical speech audiogram below 95%
- Age below 70
- Listeners should wear hearing aids on both ears for a period of at least one year

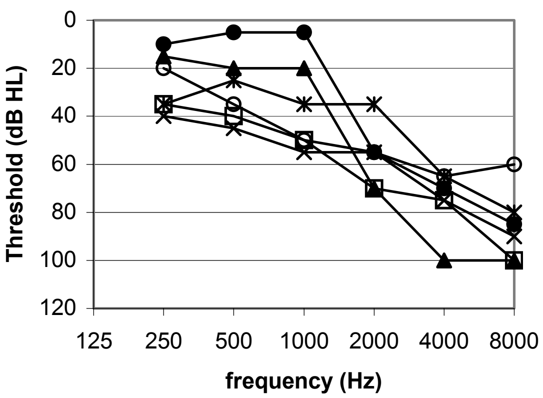


Figure 7-4 *The pure-tone audiograms of the 6 selected listeners (right ear) that completed the field experiment.*

We selected 23 patients that met these criteria, of which 7 volunteered to participate in the experiments. One of these listeners did not complete the experiments because he was annoyed by the feedback problems in the system. The pure-tone audiograms of the remaining six listeners are shown in figure 7-4 (right ears). The listeners were only paid their travel expenses.

Fitting procedure

For each individual listener the hearing-aid fitting started with three programs containing slow-acting processing only according to the settings described in the previous chapter. The balance of each of the programs was set to theoretical values according to half-gain, half-gain+5 dB and half-gain-5 dB respectively. From the results of insertion gain measurements we selected the program that gave the best resemblance to a half-gain frequency shape. This program was used as the reference condition for that individual during the experiment.
After that the analogue volume controls on the behind-the-ear hearing aids were fixed at a comfortable level for running speech at a level of 65 dB SPL. The digital volume control of the wearable unit could still be adjusted within the range of +10 and -10 dB by 2-dB steps. The same volume setting was used in all speech intelligibility experiments for a certain listener. To conclude, the three phoneme-compression programs WBC, AU and HFC were derived from the reference program by switching on the phoneme compressor in each of the three configurations.

Use of compression programs in everyday life

The three programs WBC, AU and HFC were tested pair-wise for a period of 6 weeks. Every two weeks a listener used a new possible combination of two programs. We chose not to program the reference condition for use in every-day life situations in order to force the subjects to listen with phoneme compression as much as possible. The listeners were asked to use the wearable device as much as they could. They were allowed to use their own hearing aids if necessary. Every week two compression programs were programmed as program numbers 1 and 2. Program numbers 0 and 3 were empty. The listeners did not know what combination of programs was tested. They were told that each week two new programs were programmed that could be the same or different from the previous programs. The test order of the combinations was randomized over the listeners, as well as the order of the programming on the device. Every week the listeners came to the hospital to perform speech tests and to hand in a questionnaire that they had filled in about the experience with the two programs that week.

Speech intelligibility tests

Speech intelligibility tests were performed every week. All test sessions included Consonant-Vowel-Consonant (CVC) -measurements in quiet, in multi-talker babble noise (a restaurant noise) and in single-talker noise (competing speaker played reversed). The noises were the same as used by Goedegebure et al. 2004. In addition, CVC-measurements in low-frequency noise were performed in weeks 1, 3 and 5 (car noise from Widex sound CD). The CVC-material consisted of lists of 51 items containing both meaningless and meaningful CVC-words (Steeneken et al, 1990). The words were embedded in a carrier phrase. The speech levels were set to the most comfortable level (MCL) of each individual using a loudness scaling procedure with one of the test sentences (Goedegebure et al, 2001). Signal-to-Noise-Ratios (SNR's) were determined with the reference program for each of the background noises, using an adaptive procedure with a target score of 50% phonemes correct

(Goedegebure et al, 2001). The setting of the MCL and SNR's was determined after the fitting procedure in week 0 and used in the CVC-tests of all subsequent weeks. Speech tests with sentences were performed in weeks 4 and 6, including one Speech-Reception-Threshold (SRT) measurement in quiet and one in continuous background noise of 60 dB SPL for each condition (Plomp and Mimpen, 1979).
Each week three programs were tested, consisting of the two phoneme-compression programs that the listener had been using that week and the slow-acting reference program. The order of programs in the device was randomized over the listeners. Different randomisation schemes were used for the various background-noise conditions. Before the start of a new background-noise condition four test items were presented to acquaint the subject with the new condition. In weeks 2 and 4 additional CVC- and SRT-measurements were performed with the hearing aids the listeners owned.

Experimental set-up

The listener was seated in a sound proof booth, while the experimenter was sitting outside. They had both visual and auditory contact. The experimenter typed in the responses of the subject. An extra monitor was placed in the soundproof booth to give the subjects the possibility to correct the typing of the CVC-responses. The responses were automatically saved in the computer.
The speech and the noise signals were played from two DAT-recorders and added by means of an audiometer (Madsen OB822) at any desired Signal-To-Noise-Ratio and level. The free-field output was fed into an amplifier inside the sound proof cabin and presented to the subject by a loudspeaker. A computer outside the booth controlled the audiometer and the two DAT-recorders. Almost the same configuration was used for the tests with sentences in noise. The two line outputs of a CD-player were used instead of the outputs of the DAT-recorders to feed the speech and the noise signals into the audiometer.

Questionnaire

Every week the subjects had to fill in a questionnaire before the beginning of the laboratory experiments. The questions considered the following conditions:

1. conversation with one person in a quiet environment
2. conversation with one person in a noisy environment
3. conversation with several persons
4. conversation in a street or at a shop

The listeners had to rate performance for each hearing-aid program along a so-called Visual-Analogue Scale (VAS, Aitken 1969). Afterwards the marks on a continuous line were transformed into ratings between 0 (poor performance) and 100 (good performance).

Statistical analysis

Non-parametric paired tests (Wilcoxon signed ranks test) were used for statistical testing, as the assumption of a normal distribution could not be guaranteed with the small number of subjects (N=6). SPSS version 9.0 was used for all statistical tests.

7.4 Results

Effects in time

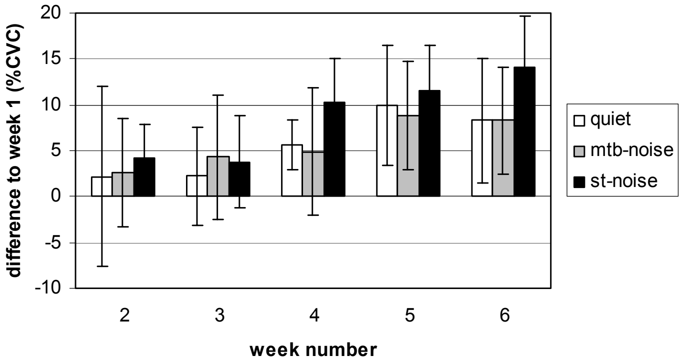


Figure 7-5 *The CVC-score with the reference program for each week number as compared to the score obtained in week number 1. The bars represent the differences in CVC-score averaged across the listeners separately for the acoustical conditions 'quiet' (light bar), multi-talker babble background noise (shaded bar) and single-talker background noise (dark bar) respectively. The error bars respresent the ±1 standard deviation.*

Figure 7-5 shows the CVC-phoneme score obtained with the *reference* program (slow AGC only) as function of week number compared to the score in week 1. The bars show the scores averaged across listeners for each of the conditions quiet, multi-talker babble and single-talker background-noise. A remarkable improvement in CVC-score can be observed with increasing week number for all conditions. The improvement found between weeks 4-6 compared to weeks 1-3 is significant for each of the three acoustical conditions (Wilcoxon, $p<0.05$). Analyses for each phoneme category separately (not shown here) show that the consonant score contribute more to the increase than the vowel score.

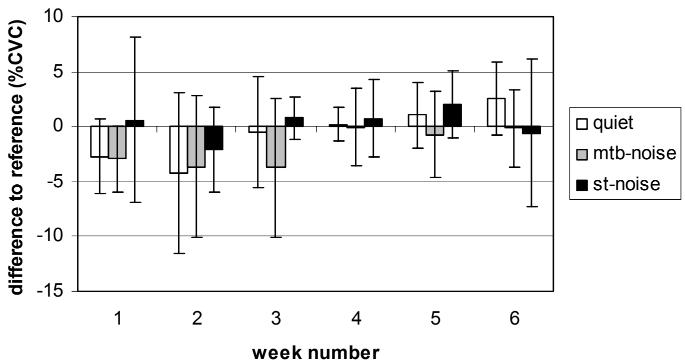


Figure 7-6 *The effect of fast compression on the CVC-score as function of week number. The bars represent the difference in CVC-score between compression programs and the reference program averaged across the listeners. The results are split for each of the acoustical conditions 'quiet' (light bar), multi-talker babble background noise (shaded bar) and single-talker background noise (dark bar) respectively. The error bars respresent the ±1 standard deviation.*

The effect of phoneme compression on the CVC-score as function of week number is presented in figure 7-6. The bars represent the differences in CVC-score between fast compression and the reference for each of the three test conditions. For each week and for each individual subject, we averaged the scores of both the used fast-compression schemes and then took the difference with the reference score. Interestingly, the effect of phoneme compression in quiet and in multi-talker babble starts to be negative, but tends to improve with increasing week number. Note that this improvement is *additional* to that found with the reference in figure 7-5. The difference between the effects of weeks 4-6 compared to those found in weeks 1-3 is statistically significant for the mtb-noise condition (Wilcoxon, $p<0.05$). A similar increase is found for both consonants and vowel.

Effects of different phoneme compression schemes

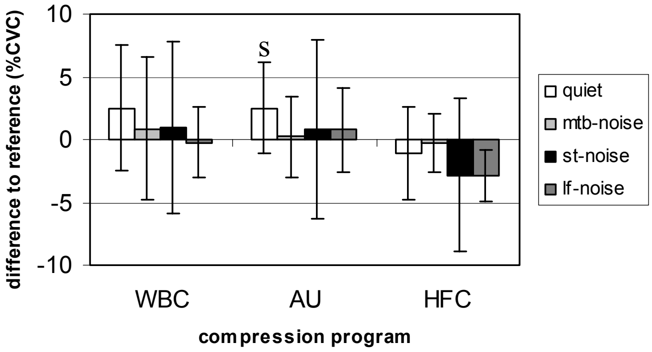


Figure 7-7 *Differences in CVC-score between the compression programs WBC, AU and HFC, and the reference program for each of the listening conditions quiet, multi-talker babble noise (mtb-noise), single-talker noise (st-noise) and 'low-frequency noise' (lf-noise)respectively. The results are averaged over all subjects in the last week the subjects tested the program. Significant effects are indicated by an S (Wilcoxon, p<0.05)*

Next we analysed the effect of the different compression schemes on speech intelligibility, taking into account a possible effect of acclimatisation as much as possible. The difference in CVC-scores with the reference has been calculated for each scheme in the last week the subject used the scheme. After that the difference scores were averaged across subjects.
Figure 7-7 shows the difference scores for each algorithm. The results are split for the four acoustical conditions quiet, multi-talker babble noise (mtb-noise), single-talker noise (st-noise) and low-frequency noise (lf-noise), respectively. The overall performance of the compression conditions is mostly better in quiet than in background noise. A significant positive effect is found for the AU-algorithm (Wilcoxon , $p<0.05$). The performance with HFC tends to be generally poorer than with the other two programs.

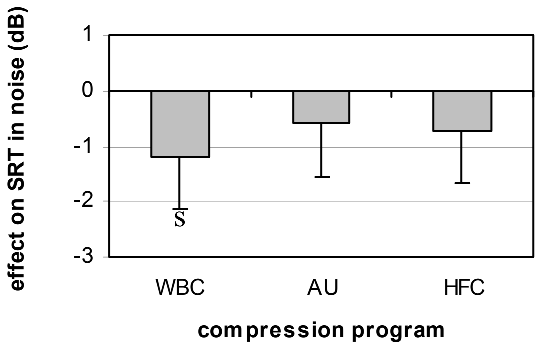


Figure 7-8 *Effect of phoneme compression on SRT in continuous noise, for each of the programs WBC, AU and HFC. A positive difference in SRT means an improvement compared to the reference program. The results are averaged over all listeners. Significant effects are indicated by an S (Wilcoxon, p<0.05).*

We also obtained Speech Recognition Thresholds (SRT's) with sentences in continuous noise. These measurements were performed only once for each condition in test week 4, so most of the possible acclimatisation effects were taken in account. Figure 7-8 shows the difference in SRT to the reference condition for each of the three algorithms. The performance with phoneme compression is generally poorer than with the reference condition (significant for the WBC-condition, Wilcoxon, $p<0.05$). Not much difference is found between the three algorithms.

Self-assessed performance in everyday environment

The performance of the subjects with the compression programs in everyday environments were assessed by means of questionnaires. Figure 7-9 shows the rating differences relative to the own hearing aid for the situations 'conversation in quiet', 'conversation in a noisy condition' and 'conversation with several persons' as function of week number. The ratings are averaged across listeners and phoneme-compression programs tested in each week. We decided to use the scores *relative to* that obtained for the own hearing aids, as these difference scores appeared to be more consistent and less sensitive to intra-individual differences than the absolute values.

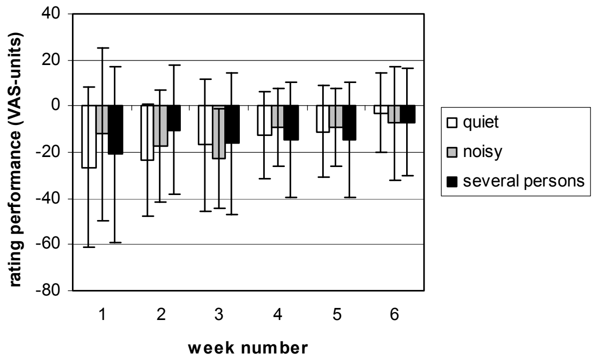


Figure 7-9 *Self-rated performance with fast compression as function of week number relative to the performance with the own hearing aids. For each week the results were averaged for the two compression programs tested in that week and the difference was obtained with the rating for the own hearing aids. The bars represent the ratings across listeners for the conditions "conversation in quiet", "conversation in noisy conditions" and "conversation with several persons" respectively. The error bars represent the ±1 standard deviation.*

In the first weeks the performance with our phoneme-compression programs is rated lower than the performance with the own hearing aids. However, in the end of the test period the difference has become negligible. The listeners tend to experience an increased benefit from the compression programs during the evaluation period, although not statistically significant (Wilcoxon, $p>0.05$). No consistent differences have been found between the pairs of phoneme-compression programs.

The experienced benefits obtained with our compression programs varied among listeners. Some listeners were mainly positive as they experienced a benefit compared to the own hearing aid, others were negative as they experienced a worse performance compared to the own hearing aids. These experienced differences in performance correlated well with the measured differences in CVC-score. Sound quality did not seem to influence much the overall judgement. The compression programs sounded generally more comfortably than the own hearing aids did. "A clearer, quieter sound" was used to indicate the difference in sound quality. Some listeners noticed that the compression programs made more speech sounds available but it seemed that they could not always make use of the information. Several times it was mentioned that the compression programs made it easier to detect a sound in background noise (e.g. hearing someone call when watching television). Although these aspects may be related to speech intelligibility in noise, it seems more a matter of distinction and sound quality than a matter of improved intelligibility. The low-level expansion was experienced as negative by most of the listeners as it worked as a kind of on-off switch in quiet environments. They complained that it was too quiet and that they missed the presence of ordinary sounds like ticking of clocks, footsteps, and the whispering of the trees. It gave no definitive solution for spontaneous feedback as it only suppresses feedback in long silent periods.

7.5 Discussion

Performance with phoneme compression

In previous studies we measured the effects of phoneme compression on speech intelligibility in a laboratory setting (Goedegebure et al, 2002a; Goedegebure et al, 2002b; Goedegebure et al, 2001). Small positive effects were found in quiet whereas predominantly negative effects were found in background noise. The main question addressed in the present study is whether regular use of phoneme compression in everyday environments would influence the performance.

In general, no large benefits of phoneme compression are found compared to the performance with slow-acting compression. Just like in previous studies, the performance with phoneme compression is generally better in quiet than in background-noise conditions. After a period of acclimatisation a small positive effect on phoneme recognition is found with one of the configurations. The CVC-recognition in background noise is not affected by phoneme compression when measured in the last couple of weeks. However, the sentence tests reveal a slightly poorer performance with phoneme compression, even after a 4-week period of using the programs. This difference in results between the sentence tests and the CVC-tests might be explained by a difference in Signal-to-Noise Ratio (SNR). The SNR-values used to perform the CVC-tests were in general more favourable, as a 50-% score with nonsense words requires a higher SNR than 50% sentence recognition. Possibly the phoneme compressor tends to deteriorate speech intelligibility mainly at a poorer SNR.

The differences in performance between the three compression conditions are not large. It is remarkable that the phoneme scores are generally poorer for condition HFC. Even in a low-frequency background noise, the performance with HFC tends to be negative. This condition is most similar to the “conventional” types of high-frequency compression. Apparently, the reduction of intensity differences within the high-frequency band does not result in the desired improvements in performance. A positive effect is found in quiet with the condition using anti-USOM processing. This was also the condition that resulted in positive effects in quiet in a previous study (Goedegebure et al. 2001).

Effects over time

Remarkable is that the performance with the programmes improves during the test period. Most of this effect can be attributed to a general improvement that is also found for the slow-acting reference condition. However, there is a tendency for an additional increase in performance with phoneme compression over time.

To explain the increase in performance with *the reference condition* we should first consider the differences in amplification with the subjects own hearing aids. Five of the six of listeners were used to hearing aids with purely linear amplification. The AGC-processing in our programs takes into account the reduced dynamic range of the impaired ear and uses less gain for moderate-to-high sound levels. This is, however, not in agreement with a study of Humes et al. 2004, who found no acclimatisation effects over a period of 6 months when evaluating two-channel AGC in a large group of hearing-impaired listeners. Another difference is the shape of the frequency response. With the test programs, including the reference condition, a steeper frequency response was obtained with more high-frequency gain and less middle-frequency gain compared to the own hearing aids. This results in a different spectral emphasis of speech cues. It is reasonable to assume that the central auditory system needs some time to adjust to this change. This may take some weeks or even months, as the reprogramming of the central auditory system is known as a relatively slow process (Gatehouse 1992). Interestingly, an increase in performance is mainly found for the consonant perception.

Next to acclimatisation to the difference in spectral shaping, also another learning effect may have played a role. By performing the speech test every week the listeners may be trained to the task of identifying and repeating the words. However, it should be noted that the speech material consisted of 20 lists containing 51 items of nonsense CVC words. This considerably reduces the chance of recognising specific words. In previous experiments we hardly found such training effects in test-retest experiments with the same speech material and one-week intervals (Goedegebure et al 2005, Goedegebure et al 2001, Verschuure et al, 1998). Therefore the contribution of this effect to the overall increase in performance is assumed to be small.

In addition to the observed improvement with the reference, the performance with phoneme compression tends to increase as well over time. Acclimatisation to the fast-acting compression seems the only plausible explanation, as the comparison to the reference condition excludes other possible learning effects mentioned in the previous paragraph. It is not surprising that the central auditory system needs some time to adapt to the processed speech. Both spectral and temporal speech cues are influenced by the fast-acting processing within the high-frequency band. The nature of these changes is probably more complex than only a shift in frequency response or overall intensity level. Effects of acclimatisation to fast-acting compression are also suggested by others (Arlinger and Billermark 1999, Toor and Verschuure 2002, Kuk et al 2003). The underlying mechanism may be a change in loudness perception due to acclimatisation as reported by (Philbert et al 2002) and (Appell et al 2002). This last study shows a stronger increase in dynamic range over time for people using fast-acting compression. This suggests that the available dynamic range of the auditory system can be reprogrammed as a response to non-linear amplification.

Considering the possibility of acclimatisation to speech processing, care should be taken to interpret the results of laboratory experiments that do not include a period of acclimatisation to the processing. In our laboratory experiments the benefits in quiet are accompanied by negative effects in background noise (Goedegebure et al 2001, Goedegebure et al 2002, Goedegebure et al 2005). The use of phoneme compression would be much more attractive if a positive effect can be obtained in quiet without causing a negative effect in background noise. The present study shows that the initial negative effects in continuous background noise tend to reduce after a period of using the system. Therefore, previous laboratory experiments may have underestimated the potential advantage of phoneme compression in current hearing-aid applications.

Experiences hearing-impaired listeners

There was no general preference for one of the phoneme compressor programs compared to the others. Although small differences were noticed between the programs, the difference between the own hearing aid dominated the overall judgement. This made it difficult to the listeners to focus to the difference between phoneme compression programs. The experienced benefit of the phoneme compression compared to the own hearing aids varied among subjects. On average no benefit was observed. Interestingly, the judgement tended to improve over time, which is in agreement with the improved CVC-scores.

Two sub-optimal choices had to be made for practical reasons. We allowed the listeners to use their own hearing aids during the trial because of the large size and limited power capacity of the experimental hearing aid. Furthermore, we did not include a slow-acting reference program for use in everyday life environments because frequent use of a slow-acting reference would slow down the acclimatisation process to the phoneme compression. Therefore, it is not possible to conclude whether the listeners prefer phoneme compression to slow-acting compression.

A remarkable observation is that the listeners experienced the sound of the phoneme-compression programs as natural and calm although we used a combination of three non-linear processing techniques. This confirms that we developed an effective solution to adapt the two-channel phoneme compressor to various speech levels as found in everyday environments. Sounds of different levels were projected in the middle of the active range of the phoneme compressor by the slow-acting compressor. A synchronised expander at the end of the system partly restored the original dynamic range. Although an extreme amount of compression was applied by the slow-acting compressor in front of the system, no annoying side effects were reported by the subjects. Therefore, the combination of synchronised compression and expansion appears to be an effective way of temporary reducing the dynamic range. This technique can be used for any speech-processing system that requires a fixed input level of speech.

7.6 Conclusions

- The phoneme score with both slow-acting and phoneme compression improves over time within a six-week period of using phoneme compression. The largest part of the effect can be contributed to an increased performance with the slow-acting compression program. A small additional increase is found for the performance with phoneme compression in multi-talker background noise. These results suggest that acclimatisation to phoneme compression may have influenced the performance.

- After a period of acclimatisation to phoneme compression, the resulting effect with phoneme compression on speech intelligibility tends to be better in quiet than in background noise. For one of the conditions a small positive effect is found on the CVC-scores in quiet, whereas a negative effect is found on the SRT in noise for another condition.
- The experienced benefit from phoneme compression varied among the listeners. It is mainly based on the difference in performance with the own hearing aids. No substantial negative experiences on sound quality were reported, indicating that we developed an effective solution to keep speech of various input levels fully within a 45-dB dynamic range of the phoneme compressor.

7.7 Acknowledgements

This research has been performed in the framework of the HEARDIP-project and the SPACE-project, both funded by the European Commission. We want to thank our European colleagues for their contribution to this study and in particular Uwe Rass from the Fachhochschule Nürnberg and Michael Stone from the University Cambridge for their great help in programming the wearable digital device. We also acknowledge the contributions made by the Heinsius Houbolt Foundation to support this research. Last but not least many thanks to the hearing-impaired listeners participating in the study for their patience and persistence needed to fulfill all experiments.

7.8 References

Appell JE, Hohmann V, Gabriel B, Kollmeier B. (2002) Evaluation of different 3-channel dynamic compression schemes in a field test with a wearable DSP prototype hearing aid. Z Audiol 41(1):22-46.

Aitken, RCB (1969) Measurement of feeling using visual analogue scales. Proceedings of the Royal Society of Medicine 62: 989-993.

Arlinger S, Gatehouse S, Bentler R, Byrne D, Cox R et al. (1996) Report of the Eriksholm workshop on auditory deprivation and acclimatization. Ear Hear 17(3 suppl):87s-98s.

Arlinger S and Billermark E (1999) One year follow-up of users of a digital hearing aid. Br J Audiol 33(4):223-232.

Bentler RA, Nelson JA (1997) Assessing release-time options in a two-channel AGC hearing aid. J Am Acad Audiol 6:43-51.

van Buuren RA, Festen JM, Houtgast T. (1999) Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. J Acoust Soc Am.105(5):2903-13.

Cox RM, Alexander GC, Taylor IM, Gray GA. (1996) Benefit acclimatization in elderly hearing aid users. J Am Acad Audiol 7(6): 428-441.

Gabriel B, Kollmeier B, Wesselkamp M (2000) Study on long-term satisfaction and on hearing aid benefit in hearing aid users. Z Audiol 39(3): 86-96.

Gatehouse S. (1992) The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. J Acoust Soc Am 92(3):1258-1268.

Gatehouse S. (1993) Role of perceptual acclimatization in the selection of frequency responses for hearing aids. J Am Acad Audiol 4(5):296-306.

Gatehouse S, Elberling C, Naylor G. (1999) Aspects of auditory ecology and psychoacoustic functions as determinants of the benefits of and candidature for non-linear processing in hearing aids. Danavox Symposium, Proceedings 18: 221-233.

Gatehouse S, Naylor G, Elberling C (2003) Benefits from hearing aids in relation to the interaction between the user and the environment. Int J Audiol. 42 Suppl 1:S77-85.

de Gennaro S, Braida LD, Durlach NI. (1986) Multichannel syllabic compression for severely impaired listeners. J Rehabil Res Dev 23:17-24.

de Gennaro S, Braida LD, Durlach NI. (1986) Multichannel syllabic compression for severely impaired listeners. J Rehabil Res Dev 23:17-24.

Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA. (2005) Evaluation of compression schemes designed to compensate for temporal and spectral masking. Int J Audiology (accepted for publication).

Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA. (2002) The effects of phonemic compression and anti-USOM on the perception of articulatory features in hearing-impaired listeners. Int J Audiol 41(7): 414-428.

Goedegebure A, Hulshof M, Verschuure J, Dreschler WA. (2001) Effects of Single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners. Audiology 40(2):10-25.

van Harten-de Bruijn HE, van Kreveld-Bos CSGM, Dreschler WA, Verschuure J. Design of two syllabic non-linear multi-channel signal processors and the results of speech tests in noise. Ear Hear 1997; 18:26-33.

Humes LE, Wilson DL (2003) An examination of changes in hearing-aid performance and benefit in the elderly over a 3-year period of hearing-aid use. J Speech Lang Hear Res 46(1): 137-145.

Humes LE, Humes LE, Wilson DL. (2004) A comparison of single-channel linear amplification and two-channel wide-dynamic-range-compression amplification by means of an independent-group design. Am J Audiol. 13(1):39-53.

Keidser G, Grant F (2001) Comparing loudness normalization (IHAFF) with speech intelligibility maximization (NAL-NL1) when implemented in a two-channel device. Ear Hear 22(6): 501-515.

Kollmeier B, Peissig J, Hohmann V. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. J Rehabil Res Dev 1993; 30:82-94.

Kuk FK, Potts L, Valente M, Lee L, Picirillo J. (2003) Evidence of acclimatization in persons with severe-to-profound hearing loss. J Am Acad Audiol 14(2): 84-99.

Marzinzik M, Hohmann V, Appel JE, Kollmeier B. (1997) Evaluation of different multi-channel dynamic compression algorithms with regard to recruitment compensation, quality and speech intelligibility. Seventh Oldenburg symposium on psychological acoustics, Oldenburg.

Moore BCJ, Glasberg BR. (1988) A comparison of four methods of implementing automatic gain control (ACG) in hearing aids. Brit J Audiol 22:93-104.

Moore BC, Peters RW, Stone MA. (1999) Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J. Acoust Soc Am 105 (1): 400-411

Munro KJ, Lutman ME (2003) The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. J Acoust Soc Am 114 (1): 484-495.

Palmer CV, Nelson CT, Lindley GA 4th (1998) The functionally and physiologically plastic adult auditory system. J Acoust Soc Am 103(4): 1705-1721.

Philbert B, Collet L, Vesson JF, Veuillet E (2002) Intensity-related performances are modified by long-term hearing aid use: a functional plasticity?. Hear Res 165(1-2): 142-151.

Plomp R, Mimpen AM. (1979) Improving the reliability of testing the speech reception threshold for sentences. Audiology 18:43-52.

Rass U (1996) A wearable signal processor system for the evaluation of digital hearing aid algorithms. Kollmeier B, editor, Psychoacoustics, speech and hearing aids. Singapore, World Scientific, ISBN 981022561X: 273-276.

Saunders GH, Cienkowski KM (1997). Acclimatization to hearing aids. Ear Hear 18(2): 129-139.

Souza PE, Bishop RD. (2000) Improving audibility with nonlinear amplification for listeners with high-frequency loss. J Am Acad Audiol 11:214-223.

Steeneken HJM, Geurtsen FWM, Agterhuis E (1990) Speech data-base for intelligibility and speech quality measurements. TNO-Institute for Perception: Soesterberg.

Toor Tv, Verschuure H. (2002) Effects of high-frequency emphasis and compression time constants on speech intelligibility. Audiology 41(7):379-394.

Turner CW, Bentler RA (1998) Does hearing aid benefit increase over time? J Acoust Soc Am 1998 103(4): 3673-3674.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP. (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Walker G, Byrne D, Dillon H. (1984) The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. J Acoust Soc Am 76(3):746-757.

Yund EW, Buckles KM. (1995a) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239.

Yund EW, Buckles KM (1995b) Discrimination of multichannel-compressed speech in noise: long-term learning in hearing-impaired subjects. Ear Hear. 16(4):417-27.

7.9 Appendix

The relations between the different input-output characteristic can be expressed quantitatively in terms of gain values. The general description of an input-output characteristic is given by:

$$y = \frac{1}{CR} x + K \qquad (7.1)$$

where y and x stand for respectively the output level and the input level, both expressed in dB. CR is the compression ratio and K is an arbitrary constant. The corresponding gain G can be defined as the difference between output and input:

$$G = y - x \qquad (7.2)$$

Substitution of (1) in (2) gives:

$$G = \left(\frac{1}{CR} - 1 \right) x + K \qquad (7.3)$$

The resulting gain G_{res} at the end of our system can be defined as the sum of the gain at the slow AGC (G_1) and the gain at the expansion system (G_3):

$$G_{res} = G_1 + G_3 \qquad (7.4)$$

An expression for the gain G_3 can be obtained by combining equations (7.3) and (7.4):

$$G_3 = \left(\frac{1}{CR_{res}} - \frac{1}{CR_1} \right) x + K_3 \tag{7.5}$$

K_3 is an arbitrary constant that defines the offset level of the input-output characteristic. For given values of CR_1 and CR_{res} the corresponding expansion factor of the third characteristic can be derived by combining equations (7.5) and (7.2) and (7.1):

$$\frac{1}{CR_3} = \frac{1}{CR_{res}} - \frac{1}{CR_1} + 1 \tag{7.6}$$

The expansion factor is defined as the reciprocal of the compression ratio. At any input x we can now calculate the expansion factor that is needed to obtain the desired resulting compression ratio CR_{res} for a given compression ratio CR_1 of the slow-acting AGC.

8 Final Discussion

The main issue in the present thesis is how speech intelligibility in hearing impaired listeners is influenced by speech processing based on phoneme compression. Of course we hoped to find improvements, but in the most critical listening conditions, noisy backgrounds, the opposite appeared to be true (chapters 4 and 6). From a clinical point of view, these negative effects can be regarded as disappointing, because the aim is to develop improved hearing aids. However, from a scientific point of view negative effects can be very informative as well. They help us to understand the influence of speech processing on speech perception under specific conditions. This knowledge is essential for a proper design of processing strategies in hearing aids. Therefore, a large part of this chapter will be dedicated to discuss the effects obtained in background noise. Before that we will focus on the positive effects found in quiet (chapter 4), as they seem to be in agreement with the rationales of our type of processing. The results of the other chapters will be used to support the discussion on technical aspects (chapters 2 and 3), perceptual aspects (chapter 5) and clinical aspects (chapter 7). To conclude we will shortly discuss the future perspectives of signal processing and compression techniques in particular within the field of audiological rehabilitation.

8.1 System design

In the general introduction we defined various rationales to apply dynamic-range compression systems in hearing aids. Obviously, our system is not designed as an optimal solution from a theoretical point of view. With its two-channel design it is not a multi-frequency processor that effectively reduces intensity differences and compensates for loudness in many frequency bands. It is based on a more pragmatic approach with speech perception as base line. A general description that covers our configurations is “a selective nonlinear enhancer of high-frequency speech cues”.

Three approaches can be distinguished that were used in our experiments:

I. increased gain in the high-frequency channel in order to emphasize cues that are important for consonant perception;
II. selective attenuation of the low frequencies in order to compensate for spectral masking (USOM);

III. reduction of intensity differences within the high-frequency channel in order to reduce temporal masking.

Table 8-1 *categorisation of test configurations with corresponding chapter numbers, sorted by type of processing*

Category number	*WBC (4,5,7) noAU (6)*	*AU (4,5,6,7)*	*AU_{HFEC} (6)*	*HFEC (4,5) HFC (4,5,6,7)*
I	+	+	-	-
II	-	+	+	-
III	-	-	+	+

Table 8-1 shows the distribution of all our configurations over the three categories. Behind the name of each configuration the corresponding chapter numbers are listed. Different approaches can be combined in one configuration. Note that categories I and III are essentially different. In chapters 2 and 3 we showed that phoneme compression with wide-band control (WBC, category I) does not reduce the dynamics of speech *within* a high-frequency band. Certain contrasts are reduced whereas other contrasts may even be increased. On average there is hardly a change in the amount of modulation or the level distribution within the high-frequency band. This is clearly a different approach than the more conventional way of compression in category III, which selectively reduces the dynamic range within the high-frequency band.

The most prominent effects were found with the anti-USOM configurations of category II. In chapter 4 we found a positive effect on initial-consonant perception and vowel perception quiet. The best results were obtained with a combination of compression with CR=2 and anti-USOM filtering. In background noise, however, remarkably negative effects were found with the same type of configuration in chapter 4 and to a lesser extent the same was found in chapter 6. The best configuration in quiet seems to give the worst results in background noise. This is one of the most intriguing results of this thesis, which will be discussed more extensively in the next sections. Before that, we may reconsider the strength and weakness in the system design.

Our compression system is not optimally designed to compensate for deficits in auditory functioning, such as loudness recruitment and loss of temporal and spectral resolution. Based on theoretical grounds the system should have a larger number of independent frequency channels. Furthermore, the systems of category I and II use the broad-band signal to control the compression within the high-frequency band. No such mechanism is present in the normal auditory system. Therefore, the system does not provide compression that should be required from a theoretical psychophysical or psychoacoustical point of view. It has been developed using a more pragmatic approach with a central role for speech perception. This may be the weakness and the strength of the system at the same time.

The relatively good performance in quiet indicates that the desired enhancement of speech cues can be obtained by using a pragmatic approach. The selective activation of the anti-USOM processing is an example of how a theoretically based compensation of USOM can be efficiently provided using a simple two-channel mechanism. It is doubtful whether similar positive effects may be obtained with a more theoretically based multi-channel system. However, in background noise the pragmatic concept seems to lose its advantages. The same mechanisms that provided benefit in quiet did not result in the desired improvements in background-noise conditions. A more theoretically based design would probably be less sensitive to changes in acoustical conditions. Normalising perceptual functions as loudness, temporal and spectral resolution makes the system less dependent of speech characteristics. Such a design would result in a general-purpose algorithm with a more similar behaviour in various kinds of acoustical conditions.

8.2 Effects on speech intelligibility in quiet

The results of chapter 5 showed that the positive effects found in quiet can be explained by an improved distinction and transmission of those features that contain mainly high-frequency information. Therefore, it seems that we achieved the enhancement of high-frequency speech cues we were aiming for. Consonant perception was improved by the phoneme compression whereas the anti-USOM processing had a positive effect on vowel perception. The signal processing seemed to do what it supposed to do. However, for a proper understanding we need more insight in the perceptual factors that may have contributed to the enhancement of high-frequency cues.

Effects on consonant perception

Focussing on consonant perception, a remarkable aspect is that no positive effect was found on the final-consonant perception. When using phoneme compression as a compensation for temporal masking we would expect to find most improvements for the final consonant, especially with the configurations of category III. The lack of such improvements implies that compensation for temporal forward masking did not seem to play a dominant role in our experiments. In literature no evidence has been found as well that particularly final consonant perception improves with phoneme compression.

Our results suggest that we should rather think in terms of a decreased vowel-consonant ratio. This is a different approach than compensation for temporal masking by compressing the intensity differences within the high-frequency band. Reducing the level difference between the consonant and the vowel does not imply that contrasts are reduced *within* the high-frequency band. The reason is that vowels always contain most energy within the low-frequencies while for consonants the energy is on average more equally distributed over low- and high frequencies. Therefore, decreasing the vowel-consonant ratio may even lead to increased contrasts within the high frequencies in favour of the consonant. The effects of decreasing the vowel-consonant ratio can be described from different perspectives. The relative amplitude difference between a consonant and a subsequent vowel within a certain frequency band can be used as a criterion to detect and identify the consonant. A change in amplitude difference between consonant and vowel can lead to disturbed perception of this cue (Hickson and Byrne 1997). This can be used as argument against decreasing the vowel-consonant ratio. Another perspective is given by Hedrick and Rice (2000), by arguing that

decreasing the vowel-consonant ratio can lead to an enhancement of spectral consonant cues, in particular for fricatives. An increased audibility of the more subtle high-frequency cues is mentioned as the underlying reason for a possible improvement. In our case the second argument seems to be applicable as we found an improved perception of frication cues. This implies that increased audibility of subtle high-frequency cues may have played a role. The frequency characteristics in chapter 4 give an important clue in this direction. Although we aimed for similar frequency responses between linear and compression conditions of category I and III, small differences can be observed in the distribution of energy within the high-frequency band. The linear condition contains relatively more energy within the 2-kHz region and less energy within the 4 kHz region relative to the compression conditions of category I and II (WBC and AU). Apparently, decreasing the vowel-consonant ratio emphasizes the speech information above 2 kHz. This confirms that an increased audibility of the subtle spectral consonant cues may have been the major reason for the positive effects found in quiet.
In addition to frication, also the transmission of plosiveness tended to improve. This is not surprising as the main part of the information for this feature is located within the high-frequency region as well. However, this feature also contains important temporal cues. It has been suggested that by using fast-acting compression these temporal cues may be negatively affected, mainly due to peaks in the onset of the phoneme (overshoots, Dreschler 1988, Verschuure et al. 1994, Moore et al. 1999). To avoid possible temporal distortion, we compensated for overshoots by using a short delay within the compression channel. Verschuure et al.1994 found a positive effect of using the same type of overshoot reduction in a similar compression system. Our results suggest that temporal cues were not distorted by the compressor, possibly because of the overshoot reduction. This is conform the results found by Souza and Turner (1996, 1998) who applied fast-acting compression in quiet conditions without noticeable distortion of temporal cues.

Effects on vowel perception

For the effects on vowel perception in quiet, a very different perceptual process must have played a role. In the anti-USOM configurations of category II the vowel energy within both low- and high-frequency band was never substantially increased. The enhancement of high-frequency vowel cues was achieved only by *removing* energy within the low-frequency region. Therefore the improved transmission of the second and third formant cues cannot be caused by an increased audibility. The only logical explanation is that the attenuation of low-frequency vowel energy causes a release of upward-spread-of-masking (USOM) for the middle and high-frequency vowel information. This was exactly the goal of the processing applied. In quiet, not many other studies have investigated effects of anti-USOM processing. Summer and Leek (1997) found positive effects on higher formant perception after at least 15 dB attenuation of the first formant. This is in line with our results. In chapter 6 we did not find a positive effect on vowel perception, most probably because the low-frequency attenuation was applied more moderately than in chapter 4. It suggests that a rigorous attenuation of low-frequency energy is needed to make this type of processing effective.

Summary

In summary, speech recognition in quiet may benefit from phoneme compression. The additional use of anti-USOM processing showed to have a complementary positive effect in our type of configuration. This means that our type of processing may help to optimize the use of the residual auditory capacities in case of a moderate-to-severe perceptive high-frequency loss.

8.3 Effects on speech intelligibility in background noise

Negative effects of anti-USOM

In background noise pronounced negative effects were found for the configurations of category II (see table 8-1) that were most successful in quiet (chapter 4). The results described in chapter 5 show that the anti-USOM processing introduced a remarkable shift in perceptual strategy that explains this paradoxical behaviour. The hearing-impaired listeners were forced to rely more on low-frequency cues as the perception of high-frequency cues was seriously disturbed by the noise. Therefore the attenuation of low-frequency energy could not be performed without negative consequences for the low-frequency speech cues, oppositely to the situation in quiet. The attenuation caused a considerable negative effect on the transmission of voicing, nasality and the first formant (see chapter 5). These are all features that are related to spectral characteristics in the low-frequency region. A negative effect of anti-USOM was even found in chapter 6, although the attenuation was applied more moderately.

In literature and clinical practise the concept of using anti-USOM is often advocated in order to achieve a possible benefit in background noise. Our results show that the opposite effect may occur as well. Can this discrepancy be explained? The spectral shape of the background noise is a key factor that may help us understand what is going on. The use of anti-USOM is often advocated because many real-life background noises have a prominent *low-frequency* character (for instance traffic noise and sounds in a reverberating environment). It is reasonable to expect that attenuation of low-frequencies helps to improve speech intelligibility in low-frequency background noise. Indeed some positive effects of adaptive low-frequency reduction have been reported for low-frequency background conditions (van Dijkhuizen et al. 1989, Rankovic et al. 1992, Fabry et al. 1993). In our laboratory experiments, however, we only used speech-like background noises. Other studies using wide-band speech-like noises did not find positive effects of anti-USOM as well (van Buuren 1995, Cook et al. 1997). Therefore our results do not exclude a possible benefit of anti-USOM in low-frequency background-noise conditions. What they do show is the possible risk of using anti-USOM in background noise conditions not dominated by low-frequency content. The general assumption that removing low-frequency energy is harmless in background-noise conditions should therefore at least be considered with the necessary caution.

Perception of high-frequency cues

It is disappointing that the negative effects of anti-USOM were not compensated by a positive effect on the perception of high-frequency cues in noisy conditions. Possibly, the rather subtle positive effects as found in quiet were not strong enough to compensate for the drastic negative effects caused by the background noise. The results of chapter 5 show that the identification of high-frequency features was seriously disturbed in background noise. The spectral masking within the high-frequency band by the noise had probably become a more relevant factor than the USOM effects. Anti-USOM processing did not affect the noise level within the high-frequency band and was therefore not effective. Furthermore, phoneme compression did not improve the audibility of fricatives or plosives as the noise level within the high-frequency bands was emphasised as well. In other words, both anti-USOM and phoneme compression were probably not successful in background noise as they did not improve the signal-to-noise ratio *within* the high frequency band. Due to the high degree of perceptual disturbance from the noise within the high-frequency band, this would have been the only possible solution to improve speech recognition.

Two important reasons can be mentioned for the high degree of perceptual disturbance at high frequencies caused by the background noise. First of all the hearing losses of our experimental population were quite severe in this frequency region (on average> 60 dB HL at frequencies of 2 kHz and above). As suggested by the model of Moore and Glasberg (1997), it has to be expected that the outer-haircell function is almost completely lost for hearing losses above 60 dB HL. This implies that the residual cochlear processing capacity of our population must have been seriously limited within the high-frequency region. A second reason is that we have used a target level of 50% phoneme score in background noise to measure the effects of signal processing. This means that for every listener we measured at a very critical signal-to-noise ratio. At less critical signal-to-noise ratios there might be a chance that the more subtle effects become more prominent, such as USOM and audibility of consonant cues. However, this is not confirmed neither rejected by literature as most evaluation tests are performed at critical signal-to-noise ratios.

Effects in a fluctuating background noise

Another disappointing result is that we did not find improvements with compression in a background noise with a fluctuating envelope (chapter 6). We hoped to find more positive effects in a fluctuating background noise than in a more stationary noise, especially when using phoneme compression schemes of category III. The concept of these configurations is to reduce temporal masking effects. Temporal masking of the speech mainly occurs when the background noise has a fluctuating envelope. Certain speech cues within the 'gaps' of the noise cannot be perceived because they are masked by the preceding noise 'burst'. It is known that hearing-impaired listeners hardly profit from the available speech information within the temporal gaps, in contrast to normal-hearing listeners. Therefore, it might be expected that phoneme compression can be used to compensate for temporal masking effects in this type of background noise. A study of Moore et al (2001) with gap detection in noise confirms that temporal-masking effects in hearing-impaired listeners caused by modulations up to 50 Hz

can indeed be reduced with fast-acting compression. Experiments with a previous version of our system suggested a positive relationship between the temporal fluctuations of the background noise and compression benefit (Verschuure et al 1998). A major difference to our experiments is hat they found the largest benefit for a background noise with a very specific modulation pattern ("printing-office noise"). Moore et al 1999 also found improvements in a background noise with a fluctuating envelope using fast-acting multi-channel compression.
However, in chapter 6 no evidence was found for a release of temporal masking by phoneme compression. There was only a tendency for improved vowel perception in the fluctuating background noise for the worst performing listeners. One possible explanation is that our type of compression is not properly designed to compensate for temporal masking. A multi-channel approach with a larger number of compression channels theoretically would result in a better reduction of modulations within separate frequency bands. Nevertheless, in chapter 2 it is shown that a substantial reduction of speech dynamics can also be obtained with the two-channel approach of category III. This means that it should be possible to compensate for temporal masking using this design.
It is possible that the use of CVC-words as test material is not applicable for evaluation in fluctuating background noise. For some words the most important speech information is located within the gaps of the noise, whereas for other words it is masked by the peaks of the noise. However, as long as this spread of speech information is random it will be averaged out for a complete list of 51 CVC-words. This is confirmed by the rather consistent scores obtained in fluctuating background noise. The only difference with the stationary noise condition was that a small learning effect was present between test and re-test. The task of repeating words in a fluctuating background noise appears to be more difficult than performing the same task in quiet or stationary noise. Continued experience and training may therefore help to improve the general performance for this condition. There are however no indications that this might have had an effect on the *difference* scores between processing conditions.
Some other studies confirm that phoneme compression does not improve speech intelligibility in a fluctuating background noise, even when using a multi-channel approach (van Harten-de Bruijn et al 1997, van Buuren et al. 1999, Olsen 2004). The experiments of Olsen et al. 2004 show interesting data, as improvements were found for normal-hearing listeners only and not for hearing-impaired listeners. This suggests that, although phoneme compression can be used to make more speech information available in a fluctuating background noise, hearing-impaired listeners cannot make use of this extra information. The possible advantage caused by a release of temporal masking seems not to play a significant role for speech recognition in hearing-impaired listeners.

Fundamental issues

Although some studies reported positive effects of fast-acting compression in background noise (Moore and Glasberg 1988, Moore et al. 1999, Verschuure et al. 1998, Yund and Buckles 1995) many others did not find clear positive effects; in some studies the effects were even negative (e.g. Bentler and Nelson 1997, van Buuren et al. 1999, de Gennaro et al. 1986, van Harten-de Bruijn et al. 1997, Kollmeier et al. 1993, Marzinzik et al. 1997, Moore et al.

2004, Vilchur 1987, Walker et al. 1984). Both the "successful" and "unsuccessful" studies included very different types of configurations, which makes it difficult to relate performance to specific characteristics of the compressor.

In chapter 6 of his book, Dillon (2001) correctly questions whether compression is an appropriate technique to improve speech intelligibility in noise. In general phoneme compression can not be considered as a noise-suppression technique because it cannot distinguish between speech and noise signals. Only when there are large differences in spectral contents between noise and speech multi-frequency compression may suppress the noise (van Dijkhuizen et al. 1989). Also in a fluctuating background noise the compressor may reduce the peak levels of the noise and increase the low-level speech cues in the gaps of the noise. In these both cases a phoneme compressor acts like a noise suppression system as it improves the signal-to-noise ratio. In the most common case of a stationary broadband background noise the speech- and noise signals are not analysed separately by a phoneme compressor. The compressor acts as an enhancer of speech(-in noise) cues. If benefits can be achieved they are based on rather subtle differences in audibility and small changes in the available speech cues. This explains that positive effects are difficult to obtain. Moreover, the small positive effects are easily counteracted by possible adverse effects.

One of these adverse effects might be the "classical" hypothesis introduced by Plomp (1989) that any reduction of speech modulations results in a reduction of available speech information. It is based on the Speech Transmission Index model (STI, Steeneken and Houtgast 1980) and experiments by Drullman et al. 1994 that show that speech intelligibility can be related to the amount of available modulations. The compressor does not only reduce the modulations of the background noise, but the speech modulations are reduced as well. From the perspective of the STI this implies a loss of speech information. Our results in chapter 3 illustrate this effect: speech compressed with a compression ratio 2 contains the same amount of modulations as speech in stationary noise at SNR=6 dB. Vilchur (1989) mentioned some strong arguments against this point of view. In the STI model modulations are reduced by adding stationary noise to the speech signal, thus the low-level speech parts are not available anymore. With compression all speech parts are still available, only the level differences are smaller. The STI model can therefore not be applied to compressed speech, which is experimentally confirmed by Hohmann and Kollmeier (1995) for compression ratios up to 3.

However, for high compression ratios there may indeed be a loss of temporal information in background noise as in some cases the performance substantially decreases as function of CR (Drullman and Smoorenburg 1997, van Buuren et al. 1999). Most probably, the amount of modulation becomes too small to be detected by the hearing-impaired listener. Our results also showed a poorer performance in background noise with CR=4 compared to CR=2. This suggests that a higher value of the CR in background noise indeed reduces the available speech information. In quiet, no such a difference in performance was found between CR=2 and CR=4. Apparently the reduction of speech modulations is less critical in quiet. The results in chapter 3 give a possible explanation for this discrepancy. The effective reduction of modulations at CR=4, expressed by the effective compression ratio, appeared to be

considerably higher for speech-in-noise than for speech alone. Therefore, the detection of relevant speech modulations will be disturbed sooner in background noise than in quiet.
Another possible adverse effect of phoneme compression is recently described by Stone et al. 2004. In their experiments on cochlear implants they found a negative effect on speech intelligibility when compressing a speech signal with one interfering speaker as background noise. No such negative effect was found when the speech and the noise were first compressed independently and added afterwards. Their explanation for the difference in performance is that it becomes more difficult to perceptually separate masker and speech when the compressor is driven by both signals at the same time. An interaction is introduced between the modulation patterns of both signals. This is what they call "co-modulation". Although they performed the experiment using a cochlear-implant simulator, the concept may also be applicable to hearing-impaired listeners. This negative effect may counteract a possible positive effect on temporal masking. In our experiments the modulation patterns of the test signal and the interfering background noise were quite different. The speech signal consisted of separate words embedded in a carrier sentence, whereas the noise consisted of a more continuously modulating running-speech signal. This may have limited the possible negative effect of co-modulation in our experiments.

Summary

In summary, no benefit of phoneme compression has been obtained in critical background-noise conditions. Poorer results are obtained when using a higher amount of compression. The use of additional anti-USOM filtering introduced considerable negative effects as well. It seems that the residual cochlear processing is too poor in our group of listeners to possibly benefit from the processing at critical signal-to-noise ratios. The attempt to provide the impaired ear with more information at high frequencies had an adverse effect on speech recognition in noise.

8.4 Influence of hearing loss

It is generally known that the shape and the amount of hearing loss can play an important role in the resulting effect of speech processing. Our type of signal processing was designed to enhance high-frequency speech cues in hearing-impaired listeners with a cochlear high-frequency loss. Therefore we focused on a group of symmetrically moderate-to-severe high-frequency losses. In all our inclusion criteria we additionally used a criterion of less than 95% maximum speech discrimination in conventional speech audiometry. By using this criterion we aimed for a group of listeners with at least some disturbance of speech recognition at normal to high levels. As a consequence we obtained a homogenous group of hearing-impaired listeners with respect to hearing loss. The advantages are that results can be compared between experiments and that averaging across listeners can be justified better than in a group with a variety of hearing losses.
We included a group of listeners with relatively high thresholds at high frequencies. This implies that the listeners had a relatively poor cochlear processing in this frequency range.

Especially the outer-haircell function may be assumed to be highly deteriorated, resulting in a reduced dynamic range and a broadening of the auditory filters. A combination of dynamic-range compression at high frequencies and anti-USOM processing is therefore a logical choice. Still, our results do not suggest that we can expect the most benefit from compression and anti-USOM for the listeners with highest thresholds and smallest dynamic ranges. Even the opposite seems to be true if we consider the results described in chapter 4. Listeners with a very small dynamic range benefited less from high-frequency compression of category III. The best results were found for patients with a dynamic range at high frequencies between 30 and 50 dB (corresponding to thresholds of about 50 to 70 dB HL). A similar inverse relationship between performance with fast-acting compression and dynamic range was found by Olsen et al. 2004. This suggests that the fitting the complete dynamic range of speech within the residual dynamic hearing range should not be the main goal of using phoneme compression. Emphasising weak phonemes seems to be a more important goal, which may provide more benefit to the listeners that still have left a reasonable residual hearing capacity at high frequencies.

Within this context it is interesting to consider the results of Turner and Cummings (1999). Their results showed that increasing audibility above 3000 Hz has no effect on speech intelligibility in quiet for listeners with severe high-frequency losses. They suggest that due to an extremely poor cochlear processing the speech information can no longer be analysed properly. In the worst case there may be no cochlear function at all. Moore et al (2000) developed a test to identify so-called “dead regions” in the cochlea. This means that there is no transmission of basilar membrane movement into neural stimulation possible at a local cochlear region, for instance due to a total loss of outer and inner hair-cell function. As a result the information corresponding to this frequency range will be delivered by neighbouring cochlear regions, which is indicated by off-frequency listening. The quality of this information is assumed to be very poor and any processing within this frequency range will probably have an adverse effect.

The presence of dead regions might be one of the explanations that we found less benefit from phoneme compression for the more severely impaired listeners (chapter 4). For instance, one of our worst performing listeners with phoneme compression was characterised by a hearing loss up to 100 dB HL at the highest frequencies and a slope of about 30 dB/octave between 1and 4 kHz. This could have been a hearing loss caused by a dead region, as they typically correspond to steeply sloping severe high-frequency losses. However, some other listeners with only slightly better thresholds at high frequency did already show benefit from phoneme compression. Therefore we do not expect that many of our listeners were characterised by hearing losses caused by dead regions. Only in some incidental cases the presence of a dead region might have influenced the performance with compression.

8.5 Phoneme compression and temporal information

An intriguing issue that remains is why no substantially improvements were found in the perception of temporal information. Most of the benefits obtained in quiet were related to improved perception of spectral information (higher vowel formants, initial fricatives), except for a tendency for an improved perception of plosiveness. In a fluctuating background noise no overall benefits were found. From a phoneme compression system we would have expected more positive effects on temporal performance. However, almost no compression system described in literature substantially improves the perception of temporal cues. On the contrary, most studies are already satisfied with the result that phoneme compression does not disturbs the perception of temporal cues in hearing impaired listeners (e.g. Souza and Turner 1996, 1998). It seems that the reason for the limited temporal effects is related to phoneme compression in general. Is there a logical explanation for the limited success of phoneme compression in the temporal domain?

It should be noted that there is a subtle but important difference between improving the perception of temporal cues and achieving a release of temporal masking. In the latter case the influence of the masker is reduced which makes some of the masked speech information available again. This information may consist of both spectral and temporal cues. Therefore, a release of temporal masking may imply an improved perception of spectral cues. This might explain the tendency for improved vowel perception in fluctuating noise for the listeners with poor speech discrimination (chapter 6).

Improving the perception of temporal information should result in improved perception of specific consonant features, especially plosiveness. Reducing intensity differences between subsequent phonemes apparently makes it not easier to identify these cues. A possible reason is that the time constants that are being used are still not short enough. For very short time constants there is a trade-off between introducing spectral and temporal distortion and being able to track the important changes in the envelope. Therefore "instantaneous" compression is not commonly used in experiments. If we find possibilities to get rid of the most important temporal distortions, for instance by using overshoot suppression like in our system, it might be possible to even increase the reaction time of the compressor. Another explanation might be to increase specific temporal intensity differences instead of reducing them, by using fast-acting expansion. It may be that the damaged cochlea needs an increased contrast to identify specific temporal cues, but the disadvantage is that temporal masking effects may increase. Some studies have tried fast-acting expansion without obtaining positive effects (Walker et al. 1984, van Buuren et al. 1999).

Another interesting point of view can be derived from the data of chapter 5. Most consonant confusions in quiet are caused by a poor transmission of features voice and place. These features contain mainly spectral information. The transmission of plosiveness is close to the ceiling of 100%. This means that there is not much to gain for the perception of temporal cues. The results of Turner et al. 1995 confirm that moderate to severe sensorineural hearing loss does not impair the temporal (nonspectral) acuity of listeners in terms of speech recognition.

8.6 Clinical implications

Impact of benefit in quiet conditions

The main reason to perform the experiments was to increase the benefit from hearing aids for an important group of users. In listening conditions without interfering noise we found small improvements, but negligible or even negative effects were found in background noise. This does not sound as a great help to the hearing-aid user, as they experience most problems with speech understanding in background noise conditions. Nevertheless, it should be realised that the small improvements found in quiet can still be meaningful. The effects are measured in comparison to a very similar reference condition in which the presentation level is optimised with regard to speech intelligibility. This means that the measured benefit can really be attributed to effects of fast-acting speech processing and not to differences in frequency-shaping or presentation level. Improving speech intelligibility over a wide range of levels can be realized additionally with slow types of compression. A second important aspect is that the improvements are obtained on top of an optimized performance under linear processing. For an overall phoneme score in quiet of 80% an improvement of 5% means a reduction of the deficit in speech discrimination by 25%. Such an improvement over a whole day of hearing-aid use may considerably reduce the listening effort. Therefore, the effect of this apparently small improvement in quiet should not be underestimated.

Impact negative effects in background noise

The price that has to be paid for an improvement in quiet seems relatively high. A reduced performance in background noise is undesirable. A possible way to benefit from the improvements in quiet without introducing problems in background noise is combining the compressor with a noise reduction scheme. In modern hearing aids combinations of compression schemes and noise reduction schemes are frequently used. An effective noise reduction algorithm can first increase the signal-to-noise ratio and after that the compression can be applied. We have evaluated such combinations of binaural noise reduction (Kollmeier et al. 1993, Wittkop et al. 1997) and fast-acting compression but this contribution did not lead to the desired effect (Goedegebure et al. 2000). A problem is that the noise reduction scheme should improve the signal-to-noise ratio considerably and this is still not possible for most general-purpose algorithms. Only with directional microphone-array techniques a substantial increase in signal-to-noise ratios can be obtained (Soede et al. 1993). The disadvantage of such algorithms is that it can only be used in a limited number of conditions and that it requires specific hardware. New algorithms using adaptive directionality may be used to overcome such problems.

A second solution would be an adaptive algorithm that automatically switches from fast-acting compression in relative quiet conditions to linear or slow-acting compression in noisy conditions. Recently the most important hearing-aid manufacturers have all introduced advanced hearing aids using some kind of environment classification. So, the technique to identify various acoustical situations is available. This allows algorithms in which the hearing-user benefits from the advantages of phoneme compression in quiet conditions without being confronted with possible negative effects in background noise conditions.

Individual fitting of speech processing

With the increasing complexity of algorithms the fitting of phoneme compression to the individual ear becomes more complex. The conventional way of fitting dynamic-range compression systems is based on loudness perception (e.g. Kiessling et al. 1996). As discussed previously, for fast-acting systems this may not be an optimum approach as it leads to a high compression ratio for the smaller dynamic ranges. This may result in adverse effects on speech intelligibility. It is therefore preferable to take a speech signal as reference signal and to use various perceptual measures of speech (e.g. quality, clarity, intelligibility) next to loudness (Moore et al. 1998, Pastoors et al. 2001). The next step is that interaction between algorithms within one hearing-aid system should also be taken in account. Franck (2004) showed that fitting procedures based on paired comparisons of speech samples can be used to find (sub-)optimal settings, but the procedures become elaborate and clinically not feasible to use.

Compression and listening comfort

Another factor that may play a role in the possible clinical use of our compression system is “listening comfort”. Until now we only considered speech intelligibility. It is often assumed that phoneme compression can affect the quality of sounds. Subjective characteristics as aggressiveness and shrill sounds are often mentioned in relation to fast-acting compression. It is interesting that none of the listeners in the experiment of chapter 7 had problems with the quality of the sound. They did hardly report any adverse experiences related to discomfort. Some reasons can be mentioned why our system is experienced as a comfortable system in spite of the fast time constants. Our system leaves the details of the short term spectrum intact by using only two channels. Hickson and Thyer (2003) confirm that speech perception is better with a two-channel compression system than using a larger number of channels. Furthermore we used a delay technique to compensate for undesired peaks resulting from phoneme compression, the so-called overshoot reduction. A last reason is that we embedded our fast-acting system in a slow-acting non-linear system for use in everyday life situations. This made it possible to use only a moderate amount of phoneme compression and thus minimises the risk for any uncomfortable side effects. Therefore, the use of phoneme compression should not necessarily coincide with substantial negative effects on listening comfort if the total system is properly designed.

Acclimatisation

The experiment of chapter 7 reveals another interesting aspect. The performance of hearing-aid users with our experimental programmes improved during the course of a field experiment. In this experiment they used the programmes in everyday environments for a period of six weeks. A causal relationship between hearing-aid use and performance after a new hearing-aid fitting was introduced as “perceptual acclimatisation” by Gatehouse (1992, 1993). The concept of perceptual acclimatisation is that the central auditory pathways need time to adjust to a structural change in neural input from the cochlea. This means that evaluation of a new hearing-aid configuration without a try-out period may underestimate the possible benefit on the long term. Acclimatisation to amplified speech has been found in several studies (Gatehouse 1992, 1993, Cox et al 1996, Munro and Lutman 2003) although others found little evidence for improvements over time (Saunders and Chienkowski 1997, Gabriel 2000, Keidser and Grant 2001, Humes and Wilson 2003, Humes et al 2004). The improvement found in our experiment can mainly be attributed to an improved performance with the reference programme. This may include a general training effect in performing the speech recognition tests. On top of that, the performance with the fast-acting compression programmes tends to improve, which can only be explained by possible acclimatisation effects. Other studies also suggest that acclimatisation plays a role in the performance with fast-acting compression (Yund and Buckles 1995b, Arlinger and Billermark 1999, Kuk et al 2003). Until there is a consensus about the size and time course of perceptual acclimatisation it should be realised that effects of speech processing may change after using the processing in everyday environments. This may have serious consequences for the interpretation of evaluation studies performed in a laboratory setting, like the experiments in chapters 4 and 6. The negative effects found in background noise should be considered with some precaution. In clinical practice the disadvantage may disappear. A larger clinical study should be performed to confirm this hypothesis.

Summary

We can conclude this section with some optimism. The clinical impact of possible benefits of phoneme compression will not be large, but can be relevant to individual hearing-aid users. The chance for negative side effects may be acceptable considering the new possibilities of signal processing and the possible occurrence of acclimatisation effects.

8.7 Future of using signal processing in hearing aids

With the present state of the art of digital techniques we can expect more complex and sophisticated algorithms to be implemented in hearing aids. Some new areas are intelligent adaptive algorithms, interactive optimisation of hearing-aid settings by users in daily environments and binaural processing techniques. Within current simple and complex algorithms, slow-acting dynamic-range compression has established an undisputable position. However, the use of phoneme compression is not trivial. In most of the current hearing aids the hearing-aid fitter can choose for different strategies, although the manufacturer defines a number of preset parameters based on the pure-tone audiogram. It is not surprising that in most cases a slow-acting strategy is preferred, both by manufacturers and hearing-aid fitters. As indicated, with the increasing use of adaptive algorithms there is a possibility to use phoneme compression selectively under specific acoustical conditions. This may help to provide certain groups of hearing-aid users with the advantages of phoneme compression without introducing undesired negative effects. A disadvantage of this approach is that due to the complexity of the fitting procedures they become unsuitable for most fitters. An alternative is setting by the manufacturer, but then the parameters are hidden for the hearing-

aid fitter. The hearing aid becomes a "black box" and the hearing-aid fitter will loose touch with the underlying speech-processing concepts. The probable result is a lack of knowledge to help those individuals that do not perform well with the average settings.
Next to the described developments, there is a tendency for using compression in a large number of independent frequency channels. This is a logical consequence of the increasing use of FFT-techniques in hearing aids. These kinds of techniques are the opposite of our approach with the use of only one active compression channel. The attractiveness of our approach are that enhancement of high-frequency cues can be achieved while keeping most of the original speech characteristics intact. More seems often better, but the question is whether the choice for many frequency channels is really motivated by performance measures or rather driven by technical reasons. Relevant spectral information may be lost.

Does the hearing-aid user really profit from all new developments? If we consider the developments of the last decennium we may conclude that the technical quality of hearing aids has considerably increased. It seems that this has mainly contributed to an improved wearing comfort of hearing aids. Dynamic compression has played a substantial role in this improvement as the sound is provided to the ear in a well-balanced manner. It is more difficult to estimate the effect of technical developments on the actual performance with hearing aids. Some limitations can be mentioned that interfere with large improvements in performance. Still, no effective general-purpose algorithm has been developed that substantially improves the signal-to-noise ratio. A major breakthrough on this topic might not be expected shortly. The most important limitation is the poor processing capacity of the damaged cochlea. Whatever hearing-aid processing is used, the processed signal has to pass the bottle-neck of a poor-functioning cochlea. The hearing aid performs only a kind of pre-processing whereas the actual processing deficits in the cochlea cannot be changed. This substantially limits the maximum benefit that can be obtained by any signal-processing application.

8.8 Methodological aspects

Within the present area of research it is difficult to achieve substantial effects in performance. If an effect can be found, it is mostly small and highly dependent on the test conditions. In our opinion it is therefore of major importance to use analysis tools that provide additional information. First of all there is a need for acoustical analysis tools as presented in chapters 2 and 3. How should we understand effects on speech intelligibility if we do not know the exact effect on the speech signal it self? Especially with the increasing complexity of signal processing the resulting acoustical effect on speech is difficult to predict. The results obtained in chapters 2 and 3 showed certain aspects that cannot be understood easily from a theoretical point of view. Secondly, the perceptual analysis methods of chapter 5 provided us with essential information about the underlying perceptual changes caused by the processing. The use of two complementary methods (INDSCAL and SINFA) proved to be very efficient in identifying the most relevant perceptual effects. Such information is not available if only sentence tests are performed, which is often the case in experimental studies.

Other relevant methodological aspects in our study are the use of a homogeneous patient group and relatively small variations in the different processing strategies. Both choices were made on purpose. It increases the chance to find consistent effects and the chance to find a possible interpretation of these effects. For a broad application it is of course interesting to know how the processing behaves for all kinds of possible hearing losses and to evaluate very distinct types of compressor configurations. The risk is however that the results become diffuse because too many variables influence the outcome. The inter-individual noise will mask the relevant effects.

Unfortunately, we were not able to perform a large clinical study to evaluate our processing schemes in everyday life situations. The experiment in chapter 7 is limited in number of subjects and design by the fact that body-worn hearing aids had to be used. It gives only an indication about what we may expect. Nevertheless, some interesting aspects are touched by the results from this chapter. As digital techniques have substantially progressed in the last couple of years, it should technically be possible by now to test the signal-processing in conventional behind-the-ear configurations. A clinical study with a larger population may provide us with relevant additional information about the possible clinical benefits.

8.9 References

Arlinger S and Billermark E (1999) One year follow-up of users of a digital hearing aid. Br J Audiol 33(4):223-232.

Bentler RA, Nelson JA (1997) Assessing release-time options in a two-channel AGC hearing aid. J Am Acad Audiol 6:43-51.

van Buuren RA, Festen JM, Pomp R (1995) Evaluation of a wide range of amplitude-frequency responses for the hearing impaired. J Speech Hear Res 38: 211-221.

van Buuren RA, Festen JM, Houtgast T. (1999) Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. J Acoust Soc Am.105(5):2903-13.

Cook JA, Bacon SP, Sammeth CA. (1997) Effect of low-frequency gain reduction on speech recognition and its relation to upward spread of masking. J Speech Hear Res 40:410-422.

Cox RM, Alexander GC, Taylor IM, Gray GA. (1996) Benefit acclimatization in elderly hearing aid users. J Am Acad Audiol 7(6): 428-441.

Dillon H. (2001) Hearing aids. Sydney: Boomerang Press.

van Dijkhuizen JN, Festen JM, Plomp R (1989) The effect of varying the amplitude-frequency response on the speech-reception threshold of sentences for hearing-impaired listeners. J Acoust Soc Am 86: 621-628.

Dreschler WA. (1988) The effects of specific compression settings on phoneme identification in hearing-impaired subjects. Scand Audiol 17:35-43.

Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am, 85(4), 1666-1675.

Drullman R, Smoorenburg GF. (1997) Audio-visual perception of compressed speech by profoundly hearing-impaired subjects. Audiology. 36(3):165-77.

Fabry DA, Leek MR, Walden BE, Cord M (1993) Do adaptive frequency response (AFR) hearing aids reduce 'upward spread' of masking? J Rehab Res 30 (3) : 318-325.

Franck BAM (2004) Hearing-aid fitting in interaction – On optimal combinations of multiple acoustical signal-processing strategies. Thesis University of Amsterdam, medical faculty.

Gabriel B, Kollmeier B, Wesselkamp M (2000) Study on long-term satisfaction and on hearing aid benefit in hearing aid users. Z Audiol 39(3): 86-96.

Gatehouse S. (1992) The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. J Acoust Soc Am 92(3):1258-1268.

Gatehouse S. (1993) Role of perceptual acclimatization in the selection of frequency responses for hearing aids. J Am Acad Audiol 4(5):296-306.

Gatehouse S, Elberling C, Naylor G. (1999) Aspects of auditory ecology and psychoacoustic functions as determinants of the benefits of and candidature for non-linear processing in hearing aids. Danavox Symposium, Proceedings 18: 221-233.

de Gennaro S, Braida LD, Durlach NI. (1986) Multichannel syllabic compression for severely impaired listeners. J Rehabil Res Dev 23:17-24.

Goedegebure A, Droogendijk JM, Verschuure J, Pastoors AD, Gebhart TM, Kiessling J, Marzinzik M, Kollmeier B (2000) Evaluation of algorithms with a combination of dynamic-range compression and noise reduction. Deliverable AC-4b SPACE project (DE3012), Telematics Application Programme EC

van Harten-de Bruijn HE, van Kreveld-Bos CSGM, Dreschler WA, Verschuure J. Design of two syllabic non-linear multi-channel signal processors and the results of speech tests in noise. Ear Hear 1997; 18:26-33.

Hedrick MS, Rice T. (2000) Effect of a single-channel wide dynamic range compression circuit on perception of stop consonant place of articulation. J Speech Lang Hear Res. 43(5):1174-84.

Hickson L, Byrne D. (1997) Consonant perception in quiet: effect of increasing the consonant-vowel ratio with compression amplification. J Am Acad Audiol. 8(5):322-32.

Hickson L, Thyer N. (2003) Acoustic analysis of speech through a hearing aid: perceptual effects of changes with two-channel compression. J Am Acad Audiol. 14(8):414-26.

Hohmann V, Kollmeier B. (1995) The effect of multichannel dynamic compression on speech intelligibility. J Acoust Soc Am. 97(2):1191-5.

Humes LE, Wilson DL (2003) An examination of changes in hearing-aid performance and benefit in the elderly over a 3-year period of hearing-aid use. J Speech Lang Hear Res 46(1): 137-145.

Humes LE, Humes LE, Wilson DL. (2004) A comparison of single-channel linear amplification and two-channel wide-dynamic-range-compression amplification by means of an independent-group design. Am J Audiol. 13(1):39-53.

Keidser G, Grant F (2001) Comparing loudness normalization (IHAFF) with speech intelligibility maximization (NAL-NL1) when implemented in a two-channel device. Ear Hear 22(6): 501-515.

Kiessling J, Schubert M, Archut A. (1996) Adaptive fitting of hearing instruments by category loudness scaling (ScalAdapt). Scand Audiol. 25(3):153-60.

Kollmeier B, Peissig J, Hohmann V. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. J Rehabil Res Dev 1993; 30:82-94.

Kuk FK, Potts L, Valente M, Lee L, Picirillo J. (2003) Evidence of acclimatization in persons with severe-to-profound hearing loss. J Am Acad Audiol 14(2): 84-99.

Marzinzik M, Hohmann V, Appel JE, Kollmeier B. (1997) Evaluation of different multi-channel dynamic compression algorithms with regard to recruitment compensation, quality and speech intelligibility. in Seventh Oldenburg symposium on psychological acoustics. Oldenburg.

Moore BCJ, Glasberg BR. (1988) A comparison of four methods of implementing automatic gain control (ACG) in hearing aids. Brit J Audiol 22:93-104.

Moore BCJ, Glasberg BR (1997) A model of loudness perception applied to cochlear hearing loss. Auditory Neuroscience 3: 289-311.

Moore BCJ, Alcantara JI, Glasberg BR. (1998) Development and evaluation of a procedure for fitting multi-channel compression hearing aids. Br J Audiol. 32(3):177-95.

Moore BCJ, Peters RW, Stone MA. (1999) Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. J Acoust Soc Am 105 (1): 400-411.

Moore BCJ, Huss M, Vickers DA, Glasberg BR, Alcantara JI (2000) A test for the diagnosis of dead regions in the cochlea. Br J Audiol, 34(4), 205-224.

Moore BCJ, Glasberg BR, Alcantara JI, Launer S, Kuehnel V. (2001) Effects of slow- and fast-acting compression on the detection of gaps in narrow band noise. Br J Audiol 35:365-374.

Moore BCJ, Stainsby TH, Alcantara JI, Kuhnel V. (2004) The effect on speech intelligibility of varying compression time constants in a digital hearing aid. Int J Audiol. 43(7):399-409.

Munro KJ, Lutman ME (2003) The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. J Acoust Soc Am 114 (1): 484-495.

Olsen HL, Olofsson A, Hagerman B (2004) The effect of presentation level and compression characteristics on sentence recognition in modulated noise, Int J Audiol 43: 283-294.

Olsen HL (2004) Supra-threshold hearing loss and wide dynamic range compression. Thesis at the Karolinska Institutet Stocholm, published by Elanders Gotab, ISBN 91-7349-921-8

Pastoors AD, Gebhart TM, Kiessling J. (2001) A fitting strategy for digital hearing aids based on loudness and sound quality. Scand Audiol Suppl. 52:60-4.

Plomp R (1989) The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function.J Acoust Soc Am 83: (2322-2327)

Rankovic CM, Freyman RL, Zurek PM. (1992) Potential benefits of adaptive frequency-gain characteristics for speech perception in noise. J Acoust Soc Am 91(1):354-362.

Saunders GH, Cienkowski KM (1997). Acclimatization to hearing aids. Ear Hear 18(2): 129-139.

Soede W, Bilsen FA, Berkhout AJ, Verschuure J (1993) Directional hearing aid based on array technology. Scand Audiol Suppl 38: 20-27.

Souza PE, Turner CW. 1996 Effect of single-channel compression on temporal speech information. J Speech Hear Res.; 39(5):901-11.

Souza PE, Turner CW. 1998 Multichannel compression, temporal cues, and audibility. J Speech Lang Hear Res. 41(2):315-26.

Steeneken HJM, Houtgast T. (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67(1):318-326.

Stone MA, Moore BCJ (2004) Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. J Acoust Soc Am. 116(4) 2311-23.

Summers V, Leek MR. (1997) Intraspeech spread of masking in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 101(5):2866-2876.

Turner CW, Souza PE, Forget LN (1995) Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J Acoust Soc Am 97 (4): 2568-76.

Turner CW, Cummings KJ. (1999) Speech audibility for listeners with high-frequency hearing loss. Am J Audiol. 8(1):47-56.

Verschuure J, Prinsen TT, Dreschler WA. (1994) The effects of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people. Ear Hear 15:13-21.

Verschuure J, Benning FJ, van Cappellen M, Dreschler WA, Boermans PP. (1998) Speech intelligibility in noise with fast compression hearing aids. Audiology 37:127-150.

Villchur E. (1987) Multiband compression processing for profound deafness. J Rehabil Res Dev 24:135-148.

Villchur (1989) Comments on " The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function". J Acoust Soc Am 86 (1): 425-427

Walker G, Byrne D, Dillon H. (1984) The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. J Acoust Soc Am 76(3):746-757.

Wittkop T, Albani S, Hohmann V, Peissig J, Woods WS, Kollmeier B (1997) Speech processing for hearing aids: noise reduction motivated by models of binaural interaction. Acta Acustica 83: 684-699.

Yund EW, Buckles KM. (1995a) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97(2):1224-1239

Yund EW, Buckles KM (1995b) Discrimination of multichannel-compressed speech in noise: long-term learning in hearing-impaired subjects. Ear Hear. 16(4):417-27.

9 Conclusions

Dynamic-range compression using short time constants (phoneme compression) is active in the range of modulations that is relevant for speech intelligibility (chapters 2 and 3). An effective reduction of intensity differences within separate frequency bands could be obtained with multi-channel phoneme compression (chapter 3). The effectiveness considerably reduced with increasing time constants (release time >30 ms). The results also depended of the test signal: the use of speech only resulted in a less effective compression than using speech in a stationary background noise. In view of the results in chapters 2 and 3, our type of compression is fast enough to change level differences between phonemes. However, subsequent intensity differences within separate frequency bands are not always reduced as required from a theoretical point of view. The processing uses a more pragmatic approach with a central role for the recognition of high-frequency speech cues.

Hearing-impaired listeners may benefit from our type of phoneme compression as we found improved speech intelligibility in quiet (chapter 4). Consonant recognition was improved by phoneme compression whereas the anti-USOM processing had an additional positive effect on vowel recognition. The positive effects found in quiet can be explained by an improved identification of high-frequency speech cues (chapter 5). The main improvements are found for the articulatory features frication (consonants) and the second/third formant (vowels). This implies that phoneme compression may help to optimally use the residual auditory capacities of the impaired ear at high frequencies.

No benefit of phoneme compression was obtained in critical background-noise conditions (chapters 4 and 6). A high amount of compression even resulted in poorer speech scores (chapter 4). Substantially negative effects were found for the configurations using anti-USOM processing (chapter 4 and 6). The hearing-impaired listeners were forced to rely more on low-frequency cues as the perception of high-frequency cues is seriously disturbed by the noise (chapter 5). Our attempt to provide the impaired ear with more information at high frequencies seemed to have an adverse effect on speech perception in noise.

Similar results were found for hearing-impaired listeners that regularly used the compression configurations in everyday environment (chapter 7). However, the general performance with phoneme compression improved over time during the evaluation period. The major part of this improvement was also found for the reference condition containing slow-acting compression only. A tendency for a small additional improvement with phoneme compression over time indicates that acclimatisation to phoneme compression may have played a significant role as well.

Summary

Hearing-aid users often continue to have problems with poor speech understanding in difficult acoustical conditions. Another generally accounted problem is that certain sounds become too loud whereas other sounds are still not audible. Dynamic range compression is a signal processing technique that may be used in hearing aids to compensate for these remaining disabilities. Its main function is to provide sufficient amplification at low input levels without overloading the auditory system at high input levels. The time constants define the time needed by the compressor to realize a change in amplification. When using relatively large time constants, the compressor only reduces differences in overall level. This type of compression is known as Automatic Gain Control (AGC) or Automatic Volume Control (AVC).With short time constants the compressor also reduces the dynamic range of a fast-fluctuating signal like speech. This last type of system is therefore often called a syllabic or a phoneme compressor. The main goal of using phoneme compression is to optimize speech intelligibility by improving the detection of weak speech cues.

We developed a phoneme compression system to improve the perception of high-frequency speech cues in hearing impaired listeners. The basic mechanism is a continuously changing balancing between low- and high-frequency amplification, steered by the input level of each speech part. As a consequence, the system should provide a relatively high amount of amplification to weak high-frequency speech cues. A specific configuration was developed to additionally reduce the negative effect of low-frequency amplification on the detection of high-frequency cues. This type of configuration is called anti-USOM processing as it is meant to compensate for "Upward-Spread-Of-Masking" (USOM) of high-frequency information by low-frequency signal parts. The main goal of the present thesis was to evaluate the effect of the different compression configurations on speech intelligibility in a group of hearing-impaired listeners with moderate-to-severe perceptive high-frequency losses (chapters 4 to 7). Additionally, we have investigated the effect of various types of compression on amplitude-modulated signals and speech(-like) signals (chapters 2 and 3).

The acoustical measurements in chapters 2 and 3 provided a good insight in the effect of compression on modulating signals like speech. Speech can be considered as a stream of sounds with a continuously varying spectrum. These spectral differences lead to fluctuations of the envelope of the signal within individual frequency bands. The modulation depth is a measure for the amount of fluctuations. Phoneme compression will normally reduce the amount of fluctuations, resulting in a smaller modulation depth. By comparing the modulation depth in a signal before and after compression, the effective amount of compression can be obtained. This method was applied using an amplitude-modulated signal (chapter 2) and

using speech(-like) signals (chapter 3). Another method compared the average level distributions of speech with and without compression (chapter 2).
The results show that relatively short time constants were needed to affect the range of modulations that are relevant to speech intelligibility. Furthermore, an effective reduction of intensity differences within separate frequency channels was only possible if the compression was applied within independent frequency channels as well. Interestingly, the results were not only influenced by the compressor settings but also by the acoustical properties of the test signal. Intensity differences were reduced more effectively for speech in a stationary background noise compared to speech only.

Chapters 4 and 6 describe the effects of different types of phoneme compression on speech intelligibility in hearing-impaired listeners. Phoneme scores were obtained in conditions with and without background noise. We evaluated the difference in performance between phoneme compression and a linear reference condition near comfortable presentation levels. This implies that the results could not be influenced by differences in overall level between the various conditions.
The results described in chapter 4 show that hearing-impaired listeners may benefit from our type of phoneme compression in conditions without background noise. Consonant perception was improved by phoneme compression whereas the anti-USOM processing had an additional positive effect on vowel perception. Unfortunately, no such positive effects were found in conditions with background noise. Even substantially negative effects were found with the anti-USOM configuration that gave the best performance in quiet. The use of a more moderate type of anti-USOM in chapter 6 also resulted in a negative effect on phoneme recognition in background noise. No benefit was found for other types of phoneme compression in background noise. The use of a compression ratio of 4 resulted even in negative effects (chapter 4). This means that the performance in background noise gets poorer with an increasing amount of phoneme compression. The temporal behaviour of the background noise did not influence the results (chapter 6). We hoped to find positive effects from phoneme compression in a fluctuating background noise, but no such improvement was found.

The results of chapter 5 can be used to understand the measured effects of compression on speech intelligibility. Two methods were used to analyse the perceptual confusions of chapter 4. INDSCAL was used to identify and visualise the most relevant differences in phoneme perception. However the interpretation of these differences was not always easy because the perceptual dimensions could be related to several perceptual features. Therefore, SINFA was used as a second method. The advantage of using this method was that the various effects could be separated for the different predefined articulatory features.
In quiet, positive effects were found on the perception of features containing mainly high-frequency information. This is according to our original goal to improve the identification of high-frequency cues by phoneme compression. However, the perception of high-frequency cues appeared to be highly deteriorated at critical background noise conditions. As a

consequence, the features containing low-frequency information had become of major importance. The use of anti-USOM processing removed low-frequency information that appeared to be relevant for the perception of low-frequency cues.

Additionally, we evaluated three phoneme compression conditions in a small field study using an experimental body-worn hearing aid (chapter 7). The phoneme compression configurations were embedded in a slow-acting non-linear system to compensate for differences in overall level. The listeners used the system for a period of six weeks next to the own hearing aids. The performance with the various compression programs was measured every week. The main question was if the performance could be influenced by a frequent use of the system.
In general, the results were similar to that in previous experiments. The performance in background noise tended to be poorer than the performance in quiet. Interestingly, the overall recognition score with phoneme compression improved over time. However, a large part of this improvement was also found for the reference condition. The tendency for a small additional improvement with phoneme compression may be attributed to acclimatization to the speech processing.
The experiences of the hearing-impaired listeners with the phoneme compression programs differed between listeners and depended of the difference in performance with the own hearing aids. Generally they had no problems with the sound of the new programs.

Samenvatting

Hoortoesteldragers houden vaak problemen met het verstaan van spraak in lastige akoestische omstandigheden. Een ander bekend probleem is dat door de versterking van het hoortoestel bepaalde geluiden als te hard ervaren worden, terwijl andere belangrijke geluiden nog steeds niet worden gehoord. Ter compensatie van deze beperkingen kan dynamiekcompressie worden toegepast in het hoortoestel. Dit is een vorm van signaalbewerking die voor voldoende versterking zorgt bij lage niveaus, zonder het gehoor te overbelasten bij hoge signaalniveaus. Een compressiesysteem heeft tijd nodig om een verandering in versterking te realiseren. Deze is instelbaar via de zogenaamde tijdsconstanten. Bij hoge waarden van de tijdsconstanten reduceert de compressor alleen globale verschillen in signaalniveau. Deze vorm van compressie wordt Automatische Volume Controle (AVC) of Automatic Gain Control (AGC) genoemd. Indien de tijdsconstanten kort gehouden worden reduceert de compressor ook intensiteitsverschillen tussen opeenvolgende spraakklanken. Deze vorm van dynamiekcompressie wordt daarom vaak aangeduid met syllabische compressie of foneemcompressie. Het doel hiervan is het optimaliseren van spraakverstaan door een verbeterde detectie van zwakkere spraakklanken.

In dit proefschrift wordt een vorm van foneemcompressie onderzocht die bedoeld is om de perceptie van hoogfrequente spraakklanken te verbeteren bij slechthorenden. Het systeem verandert continu de verhouding tussen laag- en hoogfrequente versterking, afhankelijk van de niveaus van de afzonderlijke spraakklanken. Dit leidt tot extra versterking van de zwakke klanken. Verder is een aparte configuratie ontwikkeld om de negatieve invloed van laagfrequente versterking te verminderen op de perceptie van hoogfrequente spraak ("anti-upward-spread-of-masking" of "anti-USOM" genoemd). Ons hoofddoel is om de effecten van deze vormen van foneemcompressie te evalueren op spraakverstaan bij slechthorenden (hoofdstukken 4 t/m 7). Tevens is onderzocht hoe het spraaksignaal beïnvloed wordt door verschillende vormen van compressie (hoofdstukken 2 en 3).

De akoestische metingen in hoofdstukken 2 en 3 geven een goed inzicht in hoe spraak en andere snel fluctuerende signalen beïnvloed worden door compressie. Spraak kan beschouwd worden als een continue stroom geluiden met een variërend spectrum. Deze spectrale verschillen leiden tot fluctuaties in de omhullende van het spraaksignaal binnen verschillende frequentiebanden, ook wel modulaties genoemd. De sterkte van de modulaties wordt uitgedrukt door de modulatiediepte. Foneemcompressie zal, indien effectief toegepast, de modulatiediepte binnen het spraaksignaal verkleinen. De effectieve mate van compressie kan geschat worden door de modulatiediepte voor en na compressie met elkaar te vergelijken. Deze methode is toegepast voor "eenvoudige" amplitude-gemoduleerde signalen (hoofdstuk

2) en spraak(-achtige) signalen (hoofdstuk 3). Een andere methode bestaat uit het vergelijken van de niveauverdelingen binnen spraak voor en na compressie (hoofdstuk 2).
De resultaten laten zien dat relatief korte tijdsconstanten nodig zijn om de voor het spraakverstaan relevante modulaties te beïnvloeden. Verder blijkt dat alleen compressie uitgevoerd in verschillende frequentiekanalen in staat is intensiteitsverschillen te reduceren in de afzonderlijke frequentiebanden. Opmerkelijk genoeg blijken niet alleen de instellingen van de compressor invloed te hebben op de mate van effectieve compressie, maar ook het gebruikte testsignaal. Voor spraak in stationaire ruis wordt een hogere mate van effectieve compressie gemeten dan voor spraak afzonderlijk.

In hoofdstukken 4 en 6 worden de resultaten beschreven van spraaktesten bij slechthorenden met verschillende vormen van foneemcompressie. Er zijn foneemscores bepaald in condities met en zonder achtergrondruis. Hierbij is het verschil bekeken tussen de resultaten met foneemcompressie en een geoptimaliseerde referentieconditie aangeboden op het zelfde signaalniveau.
De resultaten in hoofdstuk 4 laten zien dat slechthorenden in staat zijn te profiteren van foneemcompressie zolang er geen achtergrondgeluid wordt toegevoegd aan de spraak. Compressie blijkt vooral een positieve invloed te hebben op de herkenning van de beginmedeklinker. Daarnaast geeft anti-USOM een positief effect op de klinkerherkenning. De beste resultaten worden behaald met een combinatie van beide technieken. Helaas wordt er voor de condities met achtergrondgeluid juist negatieve effecten gevonden met deze configuratie. Ook met een meer gematigde vorm van anti-USOM blijft het resultaat negatief (hoofdstuk 6). Andere configuraties met foneemcompressie geven eveneens geen verbetering van spraakverstaan in achtergrondlawaai. Voor een grote mate van compressie wordt een verslechtering gevonden. Ook bij een fluctuerende achtergrondruis worden geen duidelijke voordelen gevonden van foneemcompressie.

De resultaten in hoofdstuk 5 geven meer inzicht in de effecten van foneemcompressie op de perceptie van spraakklanken. De verwisselingen van spraakklanken in de spraaktesten met slechthorenden zijn onderzocht door middel van twee methoden. INDSCAL identificeert en visualiseert de meest relevante verschillen in foneemperceptie door het definiëren van onafhankelijke perceptuele dimensies. De interpretatie van deze dimensies in termen van fonetische kenmerken blijkt echter niet altijd even gemakkelijk. Als aanvulling is daarom SINFA gebruikt, met als voordeel dat de verschillende fonetische kenmerken hierbij vooraf gedefinieerd kunnen worden.
Zowel foneemcompressie als anti-USOM blijken een positief effect te hebben op de perceptie van hoogfrequente spraakklanken. Met name de kenmerken fricatief en de hogere klinkerformanten worden beter doorgegeven. Dit klopt met de originele doelstelling. In de condities met achtergrondgeluid is de perceptie van hoogfrequente spraakklanken echter zodanig verstoord dat het gebruik van foneemcompressie geen uitkomst meer biedt. De laagfrequente spraakkenmerken worden nu belangrijker voor het spraakverstaan. Juist deze

blijken te worden aangetast worden door het toepassen van anti-USOM, waarmee de in hoofdstukken 4 en 6 gevonden negatieve effecten verklaard kunnen worden.

In een kleine veldstudie met experimentele digitale hoortoestellen zijn een drietal van de eerder geteste compressieconfiguraties geëvalueerd door slechthorenden (hoofdstuk 7). De foneemcompressor is opgenomen in een langzaam niet-lineair systeem waarmee de globale niveauverschillen gereduceerd kunnen worden. De programma's zijn paarsgewijs getest over een periode van 6 weken. Vanwege praktische redenen was het toegestaan om ook de eigen hoortoestellen te blijven gebruiken. Wekelijks zijn er spraaktesten uitgevoerd, met als hoofdvraag of de prestaties beïnvloed worden door een regelmatig gebruik van de programma's.
Er worden opnieuw slechtere resultaten met compressie gevonden in achtergrondlawaai. Opmerkelijk is echter dat de spraakscores toenemen over de periode van 6 weken. Een groot deel van deze toename wordt ook gevonden voor de referentieconditie met alleen langzame compressie. Het gebruik van foneemcompressie lijkt tot een lichte extra stijging te leiden. Deze kan alleen verklaard worden door gewenning aan het luisteren met foneemcompressie. De ervaringen van de slechthorenden met de compressieprogramma's zijn wisselend, afhankelijk van de prestaties ten opzichte van het eigen toestel. Het geluid wordt over het algemeen als positief ervaren.

List of abbreviations

AGC	automatic gain control
AVC	automatic volume control
AU, anti-USOM	compensation for upward-spread-of-masking
BTE	behind-the-ear (hearing aid)
CLU	categorical loudness unit
CR	compression ratio
CVC	consonant-vowel-consonant
CVR	consonant-vowel ratio
DSP	digital signal processing
FFT	fast fourier transform
FIR	finite impulse response
HCL	highest comfortable level
HFC	high-frequency control
HFEC	high-frequency emphasis control
HL	hearing level
ICRA	international collegium for rehabilitative audiology
INDSCAL	individual-difference scaling
MCL	most comfortable level
REF	reference condition
RMS	root-mean-square
SII	speech intelligibility index
SINFA	sequential information analysis
SNR	signal-to-noise-ratio
SPL	sound pressure level
SRT	speech recognition threshold
STI	speech transmission index
USOM	upward spread of masking
VAS	visual-analogue scale
WBC	wide-band-control
WDRC	wide dynamic-range compression

Dankwoord

"Papa, is je boekje al klaar?" Dat heb ik de laatste maanden regelmatig aan mogen horen. Niels vond het allemaal maar lang duren, maar voor mij klonk het als muziek in de oren. Het einde was in zicht! Dit dankzij de steun en inzet van vele mensen om mij heen.

Om te beginnen was daar de thuisbasis, het fundament.
Marieke, je hebt het vele avonden zonder mijn geestelijke aanwezigheid moeten stellen. Toch heb je me altijd onvoorwaardelijk gesteund. Aanmoedigend, meedenkend, soms ook prikkelend ("schiet dat nog een beetje op met dat hoofdstuk?"). Na een aantal erg leuke maar ook intensieve jaren is het dan zover. Het bevestigt maar weer eens hoe sterk wij samen staan.
Niels, je vrolijkheid, enthousiasme en nieuwsgierigheid is een inspiratiebron voor mij en vele anderen in je omgeving. En wie weet, kun je die eigenschappen ooit nog gebruiken om zelf een promotieonderzoek uit te voeren (je mag ook gewoon boer worden, hoor).
Mijn ouders, Jan en Jo, van jullie heb ik heel veel meegekregen. Liefde, enthousiasme, doorzettings- en relativeringsvermogen, om maar eens wat te noemen. Even terug in het Zeeuwse kon ik altijd alle drukte even achter me laten. Ik ben erg blij dat we deze nieuwe mijlpaal samen mee kunnen maken.
Mijn broer Hans Jan, je bent een rustgevende constante factor en altijd geïnteresseerd in mijn bezigheden. Vandaar ook je rol als paranimf. Hopelijk heb ik straks weer wat meer tijd om vaker een potje te tennissen, want samen op de baan zijn we natuurlijk onverslaanbaar.
Janneke en Maarten, jullie stonden altijd klaar om bij te springen als (schrijf)werk en zorg even niet meer gecombineerd konden worden. Praktische, maar vooral ook belangrijke morele steun op alle fronten. Klaas en Marc, ook jullie support was welkom.

Nu naar de kern van het proefschrift, de commissie.
Hans Verschuure, copromotor en begeleider. Als je langskwam op de hoogbouw dan gebeurde er wat: talrijke ideeën borrelden ter plekke op en de lucht wervelde nog lang nadat je vertrokken was. Je hebt me met je dynamische stijl van begin af aan gemotiveerd en geïnspireerd.
Wouter Dreschler, promotor en opleider. Je gezonde kritische blik en je gave om te structureren klinken niet alleen in dit werk door, maar ook in mijn vorming tot audioloog.
Ik wil jullie beiden bedanken voor de onmisbare bijdrage aan de inhoud van dit boekje en bovenal voor het vertrouwen dat jullie me de afgelopen jaren gegeven hebben.
Louw Feenstra, als promotor kwam je vooral in de laatste fase in beeld en vervulde daarin een prettige en daadkrachtige rol. Bedankt hiervoor.
Jürgen Kiessling, Ton van der Steen en Chris de Zeeuw, bedankt voor het commentaar en de kostbare tijd die jullie in het lezen gestoken hebben.

En dan, niet minder belangrijk, de mensen op de verschillende audiologische werkvloeren.
In de hoogbouw heerste er altijd een goede en bijzondere sfeer, vooral dankzij de boeiende mengeling van personen. Ronald Maas, behulpzaam maar vooral ook prikkelend en provocerend. Mick Metselaar, de medicus met technisch inzicht en een fijne neus voor muziek en gezelligheid. Teun van Immerseel, de filosoof en uitvinder met een goed luisterend oor. Marjon Quartel-Droogendijk, voor de nodige nuchtere en praktische inzichten, waarmee je me ook nu nog bij mag staan in de functie als paranimf. Christine ter Huurne, Thijs van Toor en Harry van Bruggen, audiologen die zich tijdens hun opleiding moeiteloos aanpasten aan de heersende hoogbouwcultuur. René de Jong, die door zijn grondige aanpak en grote inzet bijna een eeuwige student dreigde te worden. En de mensen van het kraakbeen- en neuzenlab, waar ik altijd even terecht kon als het bij "de audio" even te stilletjes werd. Bedankt allemaal voor de bijdrage aan dit werk en de onvergetelijke tijd die ik met jullie heb mee mogen maken.
Ook de band met de kliniek was heel speciaal. Met Jan, Martine, Sigrid, Elma, Paulien, Michael en de andere "AC-ers" viel er altijd wat te beleven. Uiteindelijk is dat ten goede gekomen aan zowel het patiëntgebonden stuk van dit onderzoek als aan mijn belangstelling voor de klinische audiologie.
Verder was er ook nog eens de boeiende samenwerking met de collega's binnen de projecten HEARDIP en SPACE (zie stelling 6). Met de AMC-ers Helene, Sidonne, Bas en Jan is er heel wat heen en weer gediscussieerd, zoals dat hoort tussen Rotterdammers en Amsterdammers. Ook de open sfeer in de bijeenkomsten met de andere centra liet zien dat er geen grenzen hoeven te bestaan in de wetenschap. In het bijzonder wil ik Alexandra Pastoors en Uwe Rass bedanken, na jullie werkbezoek in ons krappe flatje verliep de verdere samenwerking als vanzelfsprekend.
Na het uitvoerende werk kwam het schrijven, voor een groot deel naast mijn opleiding in het AMC. De collega's daar wil ik bedanken voor de vele leermomenten, de goede sfeer en vooral ook de steun tijdens het laatste zware jaar in Amsterdam. Terug in Rotterdam heb ik het boekje er uiteindelijk uitgeperst, daarom ook dank aan mijn huidige collega's van het Gehoor en Spraakcentrum die mij af en toe even moesten missen.

Vervolgens de vriendkring, een belangrijke randvoorwaarde in het geheel.
Ramon, Eveline, Pieter, Annemarie, Menno, Willeke en vele anderen, jullie vroegen op zijn tijd uit beleefdheid naar "het schrijven". Er volgde dan een beleefd, vaak ontwijkend antwoord van mijn kant. Nu kunnen we dat stukje overslaan en direct overgaan tot een leuker onderwerp. Jullie allen bedankt voor de steun, het geduld en de nodige afleiding. Verder vervulde de band "More about Penguins" jarenlang een belangrijke rol als gezellige, soms wat lawaaiige, maar vooral swingende uitlaatklep.

Als laatste hulde en dank aan de belangrijkste groep mensen in dit onderzoek: de slechthorenden die zich met veel inzet belangeloos onderworpen hebben aan ellenlange testen. Ik hoop dat de toekomstige hoortoestellen jullie veel goeds brengen!

Publications

Goedegebure A, Dreschler WA, Verschuure J (2005) “The effect of compression on speech modulations” submitted to Int J Audiol

Goedegebure A, Quartel-Droogendijk JM, Dreschler WA, Verschuure J (2005) “Phoneme compression in experimental hearing aids: effects of everyday-life use on speech intelligibility” in revision Int J Audiol

Goedegebure A, Goedegebure-Hulshof M, Dreschler WA, Verschuure J (2005) “Evaluation of phoneme compression schemes designed to compensate for temporal and spectral masking in background noise” accepted for publication in Int J Audiol

Goedegebure A, Goedegebure-Hulshof M, Verschuure J, Dreschler WA (2002) “The effects of phonemic compression and anti-upward-spread-of-masking (anti_USOM) on the perception of articulatory features in hearing impaired listeners” Int J. Audiol. 41(7): 414-428

Goedegebure A, Hulshof M, Maas AJJ, Verschuure J, Dreschler WA (2001). “Effects of single-channel phonemic compression on phoneme intelligibility in hearing-impaired listeners” Audiology 40(1):10-25

Goedegebure A, Droogendijk M, Verschuure J (1998). “How objective are subjects?” supplement Zeitschrift fur Audiologie: 121-123

Goedegebure A, Hulshof M, Maas AJJ, Verschuure J. (1996) “The effects of syllabic compression on speech intelligibility in hearing impaired” in “Psychoacoustics, speech and hearing aids” editor B. Kollmeier, World Scientific, Singapore pp 165-170

Goedegebure A, v.d. Steen AFW, Thijssen JM (1992) “In vitro classification of gallstones by quantitative echography”. Ultrasound in Medicine and Biology vol. 18 (6/7): 553-568

de Jager H, Goedegebure A (2003) “Het expertisecentrum Gehoor & Arbeid voor slechthorende werknemers” Tijdschrift voor bedrijfs- en verzekeringsgeneeskunde 11(1):14-17

Verschuure J, Goedegebure A, Dreschler WA (1999) “Acclimatization or learning effects in the selection of hearing aids. Z Audiol Suppl II: 198-200

Verschuure J, Maas AJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA (1996) “Technical assessment of fast compression hearing aids” in “Psychoacoustics, speech and hearing aids” editor B. Kollmeier, World Scientific, Singapore pp 171-182

Verschuure J, Maas AJJ, Stikvoort E, de Jong R, Goedegebure A, Dreschler WA (1996) “Compression and its effects on speech signal “ Ear and Hearing 17 (2): 162-175

Abstracts

Goedegebure A, Droogendijk JM, Verschuure J (2001) “Combinations of noise suppression and compression in experimental hearing aids” DGA conference Aken

Goedegebure A, Droogendijk JM, Pastoors AD, Verschuure J (2000) “Benefits of binaural hearing-aid processing in reverberant and noisy environments” XXV international congress of audiology The Hague pp 176

Goedegebure A, Verschuure J (1999) "Compression and noise-suppression in an experimental hearing aid". EFAS conference Olou

Goedegebure A, Droogendijk JM, Maas AJJ, Verschuure J (1998) "Acclimatisation of hearing-aid users to digital signal-processing". XXIV international congress of audiology Buenos Aires

Goedegebure A, Droogendijk JM, Maas AJJ, Verschuure J (1998) “Evaluatie van nieuwe hoortoestellen door middel van veldtesten”. NVA/KNO-vereniging Veldhoven

Goedegebure A, Droogendijk JM, Maas AJJ, van Bruggen H, Verschuure J (1997) “Evaluation of speech processing in everyday environments”. EFAS-conference Praag: pp 75

Goedegebure A, Hulshof M, Maas AJ, Verschuure J (1995). “The effect of syllabic compression on speech intelligibility in hearing-impaired listeners” EFAS-conference Noordwijkerhout pp 145-148

Levensloop

André Goedegebure werd op 27 december 1968 geboren in Nieuwerkerk (gemeente Duiveland). Hij behaalde in 1987 zijn VWO-diploma aan de Prof. Zeemanschool te Zierikzee. Daarna studeerde hij Technische Natuurkunde aan de Technische Universiteit Delft. Zijn afstuderen vond plaats in 1993 binnen de vakgroep akoestische perceptie, met als onderwerp het ontwikkelen en simuleren van auditieve electrofysiologische metingen. Van 1994 tot 2000 werkte hij als onderzoeker aan de afdeling KNO van de Erasmus Universiteit Rotterdam onder begeleiding van Hans Verschuure. Hier deed hij onderzoek in het kader van twee Europese onderzoeksprojecten, HEARDIP en SPACE, met als doel de ontwikkeling en evaluatie van geavanceerde hoortoesteltechnieken. In 2000 startte hij onder de hoede van Wouter Dreschler zijn opleiding tot klinisch-fysicus audioloog op de afdeling KNO van het AMC te Amsterdam. Tijdens zijn opleiding participeerde hij in een ontwikkelingsproject ter bevordering van arbocuratieve samenwerking bij slechthorendheid en was hij betrokken bij de oprichting van het expertisecentrum Gehoor&Arbeid. In 2003 rondde hij zijn opleiding af en startte hij zijn huidige functie binnen de afdeling KNO van het Erasmus MC Rotterdam. Hij is hier werkzaam als klinisch-fysicus audioloog op het Gehoor en Spraak Centrum, waarbij de kinderaudiologie binnen het Sophia Kinderziekenhuis zijn grootste aandacht heeft. Ook participeert hij actief in het Cochleair-Implant team Rotterdam, een samenwerkingsverband tussen het Erasmus MC en de koninklijke Aurisgroep.
Hij is sinds 1999 getrouwd met Marieke Goedegebure-Hulshof en heeft één zoon, Niels. Zijn huidige woonplaats is Leiden.