SIMULATION OF HEARING IMPAIRMENT BASED ON THE FOURIER TIME TRANSFORMATION

Josef Chalupper and Hugo Fastl

Institute for Human-Machine Communication, Technical University of Munich
80290 Munich, Germany, {cha, fas}@mmk.ei.tum.de

ABSTRACT

A signal processing system was designed to simulate hearing impairment for normal hearing listeners. Typical consequences of sensorineural hearing loss are an elevated threshold in quiet, abnormal growth of loudness ('recruitment') and reduced spectral and temporal resolution. However, the just noticable difference (jnd) in intensity is unaffected. Based on an aurally adequate time-frequency representation, these effects can be simulated by intensity quantization and frequency dependent dynamic level expansion. The Fourier Time Transformation (FTT) was used for analysis of the time signal and the inverse FTT for resynthesis of the modified complex spectra. Input-output functions used for level expansion are calculated from the audiogram and a categorical loudness scaling procedure. Hearing experiments were carried out to find the appropriate quantization step and to simulate intensity jnd's of normal and hearing impaired listeners. The proposed processing scheme will be used to improve the design and fitting of hearing aids.

1. INTRODUCTION

Besides providing normal hearing listeners with realistic demonstrations of auditory consequences of hearing impairment - which may be very helpful in the prevention of hearing disorders -, an accurate simulation of hearing impairment can be of great value for the development of signal processing algorithms in future hearing aids. By restricting the simulation to certain aspects of hearing loss, it is possible to gain insight how these aspects combine and result in reduced speech intelligibility in noise ('cocktail-party effect'). A deeper understanding of the processes involved in speech perception should lead to improved signal processing algorithms and fitting procedures for hearing aids. New developments can be evaluated by simulation with normal hearing listeners, which gives the possibility to test many subjects with the same - simulated - hearing impairment and thus - with statistical procedures - to rule out individual 'bias' effects. Without simulation, usually each hearing impaired subject has a different hearing loss, and therefore bias effects often make it difficult to analyze psychoacoustic data and relate them to certain aspects of hearing impairment or signal processing. Important psychoacoustic findings, a simulation of hearing impairment should account for, are:

- elevated threshold in quiet
- abnormal growth of loudness
- reduced spectral and temporal resolution
- unaffected intensity jnd.

The most obvious perceptual consequence of cochlear damage is the elevated threshold in quiet, which can be determined by the audiogram. Above threshold, increasing level by the same amount in normal and hearing impaired listeners, often leads to a larger increment of loudness in hearing impaired listeners. This behaviour usually is called 'recruitment' [1]. Reduced spectral and temporal resolution can be shown in masking patterns obtained with hearing impaired subjects [2]. In contrast, however, intensity resolution seems - to a first approximation - not to be affected by cochlear damage [3]. Unfortunately, this means for hearing impaired people with recruiting ears that, because they have a reduced dynamic range (between threshold in quiet and threshold of pain), there is a smaller number of intensity jnd's left. Since this may result in a reduced ability to distinguish different sounds, it also may contribute to the 'cocktail-party effect'.

There are two major classes of hearing loss simulations that have been studied to date. The first one consists of simulations using additive noise to elevate threshold in quiet [4]. Disadvantages of this method are (1) that recruitment is mediated by partially masked loudness (caused by the additive noise) and therefore can not be studied separately from elevated threshold; and (2) simulating severe hearing losses may cause cochlear damage to normal hearing subjects. The second class are processing schemes based on multiband dynamic expansion [5]. To account for reduced spectral and temporal resolution, additionally different techniques for spectral [6] and temporal [7] smearing are applied.

A common shortcoming of all these simulations is that intensity jnd's are decreased as a direct consequence of level expansion. This means to improve intensity resolution, which is in contrast to psychoacoustic data. Another problem is that often either no aurally adequate time-frequency representation is applied (e.g. short term FFT) or the temporal fine structure (phase, time alignment) is not preserved.

Therefore, a processing scheme is proposed, that (a) can simulate the above mentioned important perceptual consequences of hearing impairment and (b) overcomes the shortcomings of other simulation methods.

2. METHOD OF SIMULATION

The first stage of the proposed signal processing system, as presented in Fig. 1, is to convert the time signal into an aurally adequate time-frequency representation.
The requirements of such a representation are:

- quasi-continuous course in both frequency- and time-domains.
- temporal and spectral resolution close to those of the human auditory system.
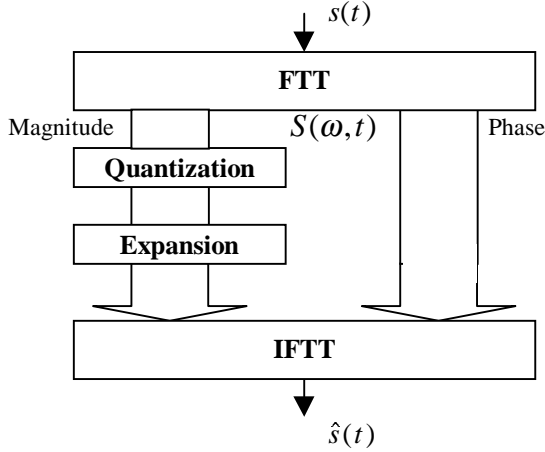- perceptually perfect reconstruction.

Figure 1. Block diagram of the sequence of processing stages for simulating hearing impairment.

Since the bandwidth of the auditory filter expressed in critical bands [2] is nearly frequency independent below 500 Hz but increases with frequency above 500 Hz, neither short-term-Fourier-analysis (frequency resolution constant) nor wavelet transform (resolution proportional to frequency) can fulfill these requirements. In theory, quasi-continous time-frequency analysis with arbitrary bandwidths is possible with auditory filterbanks [8] or the Fourier Time Transformation [9]. Since perceptually perfect reconstruction with the FTT is practically available from [10], spectro-temporal analysis/synthesis is done by FTT/IFTT.

In the next stages (Quantization and Expansion), only magnitudes of the complex FTT-spectrum are processed. In order to simulate unaffected intensity resolution despite abnormal growth of loudness, magnitudes are first quantized and expanded thereafter. After recombining modified magnitudes and orignal phases, a time signal is resynthesized by inverse FTT.

The proposed method for simulation of hearing impairment should be able to account for the above mentioned perceptual consequences of cochlear hearing loss. Abnormal growth of loudness, elevated threshold in quiet, and unaffected intensity jnd are directly 'implemented' in the processing stages 'Quan-tization' and 'Expansion'. Reduced frequency selectivity in terms of flattened tuning curves is a consequence of dynamic expansion [5]. The same is true for reduced temporal resolution, because nonsimultaneous masking for hearing impaired subjects can be modeled by a - compared to normal hearing - 'more' expansive nonlinearity followed by an unmodified temporal window [11].

By simply jittering amplitudes on the time or frequency axis, it is possible to study the effects of reduced spectral and temporal resolution separately.

2.1 Analysis/synthesis by FTT

The Fourier Time Transformation (FTT) of a causal signal s(t) was defined by Terhardt [9] as

$$S(\omega,t) = \int_0^t s(\tau)h(\omega,t-\tau)e^{-j\omega t}d\tau, \qquad (1)$$

where $h(\omega,t)$ is a frequency-dependent temporal window [10]:

$$h(\omega,t) = 2a(\omega)\frac{a(\omega)^{n-1}t^{n-1}}{(n-1)!}e^{-a(\omega)t} \qquad (2)$$

with

$$a(\omega) = \pi\frac{B_{3dB}(\omega)}{\sqrt{2^{1/n}-1}}. \qquad (3)$$

To achieve an aurally adequate frequency resolution, $B_{3dB}(\omega)$ is chosen proportionally to the critical bandwidth Δf_G of the auditory system [2]:

$$\Delta f_G = \left\{25 + 75\left[1+1,4\left(\frac{f}{kHz}\right)^2\right]^{0,69}\right\}Hz \equiv 1Bark \qquad (4)$$

When applying a frequency-independent window (e.g. Hanning-window) the short-term-Fourier-transform [12] of s(t) would be obtained. If the order n of $h(\omega,t)$ is set to n = 4, the resulting window has the shape of the impulse response of a 'gammatone filter' [13].

The inverse FTT [10] is given by

$$\hat{s}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(S(\omega,t)*x(\omega,t))e^{j\omega t}d\omega, \qquad (5)$$

where $x(\omega,t)$ is a frequency-dependent correction to ensure that across frequency, temporal windows - despite their frequency dependent shape - reach the same maximum height at the same time with the same phase.

For a fixed analysis frequency ω_A, the FTT can be depicted as a complex modulator followed by a convolution with $h(\omega,t)$:

$$S_{\omega_A}(t) = S(\omega,t)|_{\omega=\omega_A} = (s(t)e^{-j\omega_A t})*h_{\omega_A}(t) \qquad (6)$$

Thus, replacing the integral in (5) by a finite number of frequencies yields

$$\hat{s}(t) = \sum_{\substack{\omega_A=\pm\omega_{A_i}}} [(s(t)e^{-j\omega_A t})*h_{\omega_A}(t)*x_{\omega_A}(t)]e^{j\omega_A t}\frac{\Delta\omega_A(\omega_A)}{2\pi}, \quad (7)$$

which can be interpreted as a bank of overlapping filters with center frequencies ω_A and bandwidths $\Delta\omega_A$. Since the signal in one channel of this filterbank can be viewed as bandlimited, the (channel-)sampling rate for a time discrete implementation should not fall below $\Delta\omega/\pi$. Signal reconstruction from downsampled and modified channels require a synthesis window to suppress distortions outside the channel bandwidth. Thus, a synthesis window is introduced by substituting the window in (7) by a series connection of an analysis window and a synthesis window:

$$h_{\omega_A}(t) \rightarrow h_{\omega_A}^{A*S}(t) = h_{\omega_A}^A(t) * h_{\omega_A}^S(t) \qquad (8)$$

In a similar manner the correction $x(\omega,t)$ in (5) is split up in an analysis and a synthesis correction. The resulting system is implemented as time aligned analysis/synthesis filterbank. Figure 2 shows the sequence of processing in one channel, where $S_{corr}(\omega_A,t)$ denotes the corrected complex FTT Spectrum.
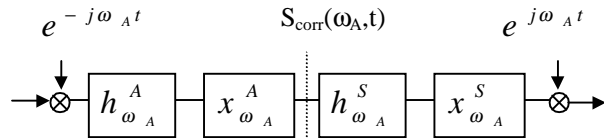
Figure 2. Implementation of FTT-based analysis/synthesis

For the current application, a window of order n=4 is chosen with a bandwidth B_{3dB} of $0.887\Delta f_G$ to create filters with an equivalent rectangular bandwidth (in intensity) of 1 Bark. The channel sampling rate is set to 7350 Hz, which corresponds to $\Delta\omega/\pi$ in the broadest filter used. Analysis frequencies are spaced with 0.2 Bark in a range from 20 Hz to 16 kHz, to obtain a perceptually perfect reconstruction. For further details concerning aurally adequate parameter setting of the FTT see [10].

2.2 Quantization

Modulation thresholds of slowly amplitude modulated sounds can be calculated from the excitation pattern model for intensity discrimination originally presented by Zwicker and Maiwald (see [2]). They assumed that the threshold is reached when the excitation level changes somewhere along the critical-band rate scale by more than 1 dB. Since the upper slope of a sinusoids excitation pattern grows more rapidly than stimulus level, their model also accounts for the modulation detection of high-level sinusoids for which a change in stimulus level less than 1 dB is sufficient. Therefore, the magnitudes of $S_{corr}(\omega_A,t)$ are level and frequency independent quantized with a step size of 1 dB to simulate unaffected intensity jnd's in hearing impaired listeners. Inevitable distortions introduced by nonlinear processing like quantization and expansion are attenuated - but not entirely eliminated - by the synthesis window.

2.3 Expansion

Input-output functions for simulating a specific hearing impaired listener are derived in two steps:
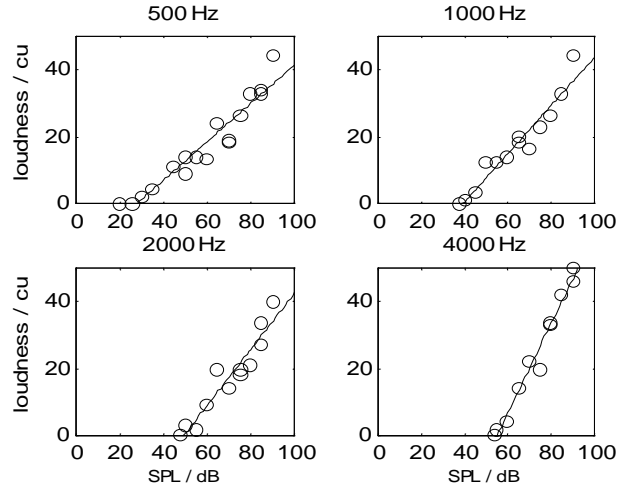
Figure 3. Results of categorical loudness scaling.

(1) The necessary level to yield the same loudness for normal and impaired hearing at four frequencies is determined from the results of a categorical loudness scaling procedure called 'Würzburger Hörfeld' (WHF) [14]. Fig. 3 shows the results for a hearing impaired listener with a sloping hearing loss (o) and a line fitted to the data.

(2) The resulting input-output functions are interpolated across frequency with regard to the course of the threshold in quiet as measured by the audiogram. As an example, the function used for dynamic level expansion at 4 kHz is depicted in Fig. 4. as calculated from data shown in Fig. 3.
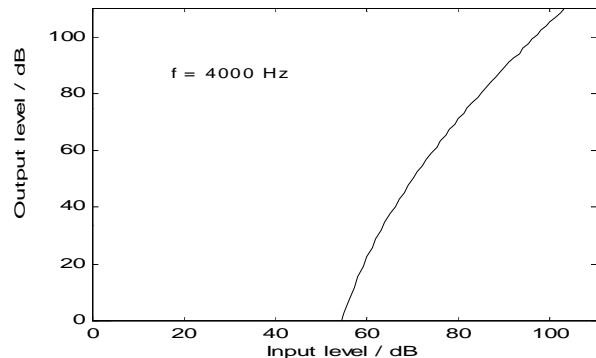
Figure 4. Expansive input-output function at 4000 Hz

3. EXPERIMENTAL RESULTS

To evaluate the proposed method for simulating hearing impairment and to check the quantization step of 1 dB, thresholds for detection of amplitude modulation were measured for simulated normal and hearing impaired subjects as a function of quantization step. A slowly modulated sinusoid (carrier frequency 4070 Hz, modulation frequency 4 Hz, level 80 dB SPL) was used as stimulus. Thresholds were determined by
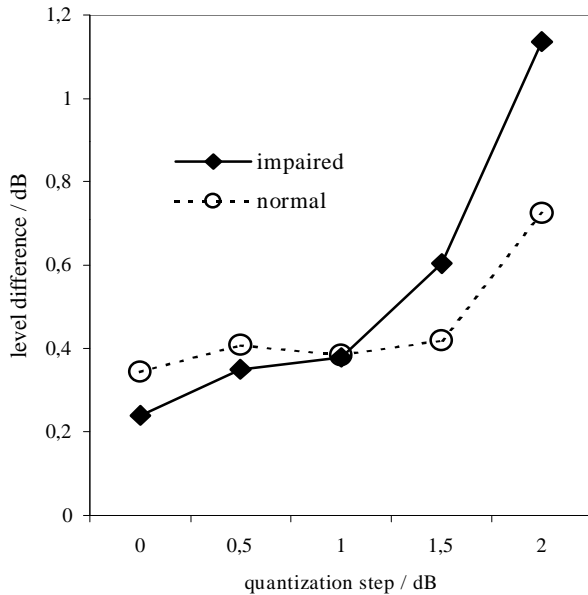
Figure 5. Modulation thresholds of simulated normal and hearing impaired subjects for different quantization steps.

a 2AFC procedure with constant stimuli. To simulate normal hearing, no level expansion was applied. As a typical example for presbyacusis, hearing impairment was simulated based on audiological data of the person, whose WHF is shown in Fig. 3. Mean results of four subjects are presented in Fig. 5. Modulation thresholds for normal hearing are for quantization steps less than 2 dB at about 0.4 dB, which is in accordance with data from the literature [2]. As expected, with no quantization, the threshold for simulated hearing impairment is lower by a factor, which corresponds to the slope of the expansive function in Fig.4. For a quantization step of 1 dB, thresholds for normal and impaired hearing are equal, whereas with increasing quantization, impaired thresholds are higher than for normal hearing. Thus, to simulate unaffected intensity resolution, a quantization with 1 dB seems to be appropriate. For a quantization step of 2 dB, strong distortions are introduced, so that even the originally unmodulated standard gets modulated. Since the difference limen of fluctuation strength grows with modulation depth [15], thresholds at 2 dB are larger for the simulated hearing impaired.

4. SUMMARY

A signal processing scheme for simulating hearing impairment based on the Fourier Time Transformation is proposed, which overcomes the shortcomings of former simulation methods. Besides elevated threshold in quiet, abnormal growth of loudness, reduced temporal and spectral resolution, for the first time also unaffected intensity resolution can be simulated, which is shown by measuring modulation thresholds for simulated normal and impaired hearing.

The system allows new signal processing algorithms and fitting procedures for hearing aids to be evaluated, especially with regard to speech intelligibility in noise.

REFERENCES

[1] Steinberg, J., Gardner, M., "Dependence of hearing impairment on sound intensity". J. Acoust. Soc. Am. 9, 11-23, 1937.

[2] Zwicker, E., Fastl, H., "Psychoacoustics". Springer-Verlag (Berlin Heidelberg New York) ,1999.

[3] Florentine, M., Reed, C.M., Rabinowitz, W.M., Braida, L.D., Durlach, N.I., Buus, S., "Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss". J. Acoust. Soc. Am. 94, 2575-2586, 1993.

[4] Florentine, M., Fastl, H., Buus, S., "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking". J. Acoust. Soc. Am. 84, 195-203, 1988.

[5] Lum, D.S., Braida L.D., "DSP Implementation of a real-time hearing loss simulator based on dynamic expansion". In: Modeling Sensorineural Hearing Loss, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), 1997.

[6] Nejime, Y., Moore, B.C.J., "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise". J. Acoust. Soc. Am. 102, 603-615, 1997.

[7] Drullman, R., Festen, J.M., Plomp, R., "Effect of temporal envelope smearing on speech reception". J. Acoust. Soc. Am. 95, 1053-1064, 1994.

[8] Irino, T., Unoki, M., "A time-varying, analysis/synthesis auditory filterbank using the gammachirp". Proc. ICASSP98, 6, 3653-3656, 1998.

[9] Terhardt, E., "Fourier transformation of time signals: conceptual revision". Acustica 57, 242-256, 1985.

[10] Mummert, M., "Speech coding by contourizing an aurally adapted spectrogram and its application to data reduction". In German. Ph.D. thesis, Technische Universität München, 1997. For an English description and software download see also http://home.t-online.de/home/Markus.Mummert

[11] Oxenham, A., Moore, B.C.J., "Additivity of masking in normally hearing and hearing-impaired subjects". J. Acoust. Soc. Am. 98, 1921-1935, 1995.

[12] Allen, J.B., "Short term spectral analysis, synthesis and modification by Discrete Fourier Transform". *IEEE Trans. ASSP*, vol. ASSP-25, 235-238, June 1977.

[13] Patterson, R.D., Holdsworth, J., Nimmo-Smith, I., Rice, P., SVOS Final Report: The Auditory Filterbank. APU report 2341, MRC Applied Psychology Unit, Cambridge, UK.

[14] Heller, O., "Orientated category scaling of loudness and speech-audiometric validation". In: Schick, A. (ed.), Contributions to Psychological Acoustics, Vol. 5, Oldenburg, BIS, 135-159, 1991.

[15] Schöne, P., "Fluctuation strength of Amplitude-Modulated Tones". Acustica 41, 252-157, 1979.