Statistical representation of sound textures in the impaired auditory system

RICHARD MCWALTER* AND TORSTEN DAU

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

Many challenges exist when it comes to understanding and compensating for hearing impairment. Traditional methods, such as pure tone audiometry and speech intelligibility tests, offer insight into the deficiencies of a hearing-impaired listener, but can only partially reveal the mechanisms that underlie the hearing loss. An alternative approach is to investigate the statistical representation of sounds for hearing-impaired listeners along the auditory pathway. Using models of the auditory periphery and sound synthesis, we aimed to probe hearing impaired perception for sound textures – temporally homogenous sounds such as rain, birds, or fire. It has been suggested that sound texture perception is mediated by time-averaged statistics measured from early auditory representations (McDermott *et al.*, 2013). Changes to early auditory processing, such as broader "peripheral" filters or reduced compression, alter the statistical representation of sound textures. We show that these changes in the statistical representation are reflected in perception, where listeners can discriminate between synthetic textures generated from normal and impaired models of the auditory periphery. Further, a simple compensation strategy was investigated to recover the perceptual qualities of a synthetic sound texture generated from an impaired model.

INTRODUCTION

The healthy auditory system is capable of processing many sounds with varying spectral and temporal features. These sounds range from the simplest artificial stimuli, such as a tone, to the most complex auditory scene, composed of such elements as the "cocktail party", music, or environmental sounds. A sensorineural hearing-impaired system, on the other hand, demonstrates weakness in processing almost all sounds as compared to the normal, healthy ear. The simple artificial tones are no longer audible for particular levels and frequencies. The auditory scene becomes overwhelming as the attention-driven source separation is no longer able to track the target sound. These changes are mostly attributed to the degradation of early auditory processing, such as broadening of "peripheral" filters and loss of compression, which in turn modifies the representation of sounds at higher stages of the auditory system.

Although environmental sounds have been used in speech-in-noise experiments, their processing and perception remains relatively unstudied in the impaired auditory system. Investigating the perception of environmental sounds in the impaired auditory

*Corresponding author: rmcw@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

system could prove beneficial for understanding the difficulties such listeners have in complex listening environments. One possible avenue is to explore the representation of sound textures – temporally homogeneous sounds such as rain, birds chirping or fire – that are composed of the superposition of many similar acoustic events. It has been shown that the perceptual qualities of sound textures can be captured using a standard model of the auditory system and a set of *texture* statistics (McDermott and Simoncelli, 2011).

In this study, we investigated the auditory systems' sensitivity to synthetic sound textures generated with various impaired models of the auditory periphery. Using normal-hearing listeners we probed the response to two major factors in sensorineural hearing loss; broader peripheral filters and loss of compression. In addition, we quantified the effects of the impaired synthetic textures by parametrically varying the synthesis system statistics. Lastly, we developed a compensation strategy to optimize the texture statistics in an attempt to regain the perceptual qualities of sounds generated from impaired models towards that of an original texture.

SOUND TEXTURE ANALYSIS AND SYNTHESIS

The generation of sound textures can be accomplished by *shaping* Gaussian noise with original sound texture statistics measured from a standard model of the auditory system (McDermott and Simoncelli, 2011). The model accounts for fundamental spectral and temporal processing by using a set of cascaded filter banks. The *texture* statistics are measured on the envelope of a filtered original sound texture, which capture the time-averaged envelope distributions as well as the covariance between pairs of neighboring filterbank channels. A companion synthesis component accepts the statistics and modifies a Gaussian noise signal, such that the statistics of the original sound texture are imposed on the synthetic sound. The synthesis process facilitates the exploration of the model structure and the statistical parameters to investigate the change in texture representation and their consequences on perception.

The auditory model is composed of three main components; peripheral frequency filtering, compression and envelope extraction, and modulation filtering as shown in Fig. 1: Analysis System. The peripheral filtering is accomplished by means of a gammatone filterbank, where the normal-hearing system uses equivalent rectangular bandwidth (ERB) spaced filters (Glasberg and Moore, 1990). A power-law compression is applied to the output of each peripheral filter signal followed by computing the absolute value of the discrete time analytic signal, resulting in the subband envelope (Harte *et al.*, 2005). The final stage is a modulation filterbank, which is composed of octave-spaced bandpass filters (Dau *et al.*, 1997).

Statistics that capture many perceptually significant features of sound textures have been identified by McDermott and Simoncelli (2011). These include marginal moments and pair-wise correlations, measured on the envelope signals of the peripheral filters and modulation filters. The envelope signals are down-sampled to 400 Hz at the output of the peripheral filter, as shown in Fig. 1: Synthesis System. The statistics

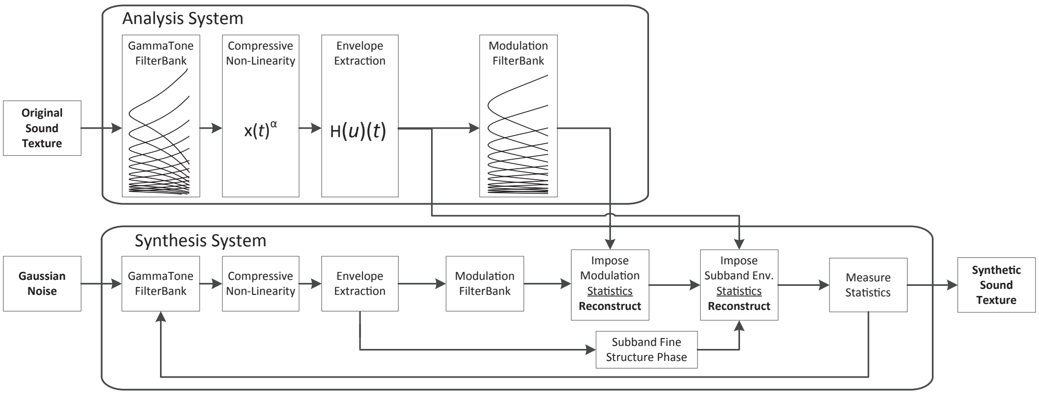

Fig. 1: Implementation of the texture synthesis system (McDermott and Simoncelli, 2011). The system is comprised of an auditory inspired analysis component, which measures marginal moments and pair-wise correlations. The statistics are passed to the synthesis component, which imposes the *texture* statistics on a noise input.

can be grouped into two main categories; the subband envelope statistics and the modulation statistics. The subband envelope statistics include marginal moments (mean, coefficient of variance, skewness, and kurtosis) and pair-wise correlations measured across the eight neighboring subbands. The modulation statistics include the modulation power measured at the output of each modulation filter, as well as pair-wise correlations measured for a specific modulation filter center frequency across the neighboring peripheral subbands.

The synthesis of sound textures is accomplished by imposing the statistics measured from the auditory model (Analysis System) to a Gaussian noise input. The synthesis system operates in two domains; the subband envelope and modulation domain. The synthesis system begins by deconstructing the noise signal to the modulation domain and applying both the modulation power statistics and modulation correlation statistics. The modulation filtered signals are then reconstructed to the subband envelope form, where the marginal moments and pair-wise correlation statistics are imposed. The subband envelope signals are then recombined with the subband fine structure phase signal and reconstructed to the time-domain signal.

Synthetic textures were generated to functionally account for changes to the auditory system caused by sensorineural hearing loss. The limited frequency selectivity is modeled by broadening the peripheral gammatone filters and the loss of compression is modeled as an increase in the power-law compression (Moore, 2007; Rosengard *et al.*, 2005). Figures 2A and 2B show the filter bandwidth and compression ratio used to generate the synthetic textures. The cross-over level for neighboring filters was preserved in all models, which resulted in fewer peripheral filters being used for the impaired hearing model. In turn, this reduced number of peripheral filters reduces the number of parameters measured for each textures. A comparison of the peripheral filterbank structure is shown in Fig. 2C.

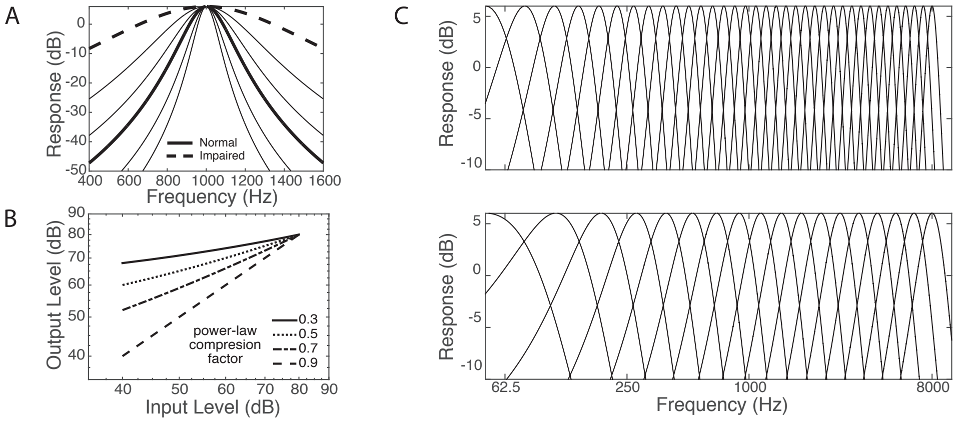

Fig. 2: Comparison of normal and impaired model configurations. (A) Simulated peripheral filter bandwidth for normal and impaired (4×) listeners. (B) Power-law compression ratio input-output level between normal ($\alpha = 0.3$) and impaired ($\alpha = 0.9$). (C) Filterbank model of frequency selectivity for normal (upper) and impaired (lower) hearing.

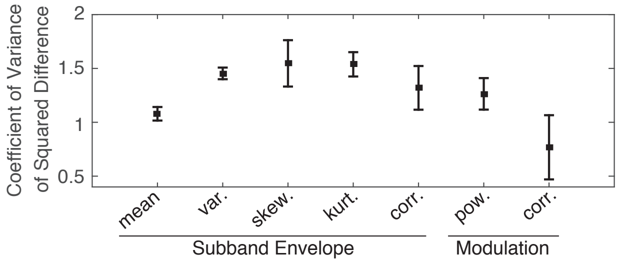

Fig. 3: Normalized coefficient of variance comparing the normal and impaired synthetic texture statistics.

Textures synthesized with impaired models of the auditory periphery alter the representation of the sound textures, as shown in Fig. 3. In order to characterize this change, we generated 45 different textures with a normal and an impaired model with four times broader filters. The textures, including birds chirping, babble, river flowing, and jackhammer, were selected to span the space of statistics, and therefore also covered a broad range of perception. The synthetic sounds were then analyzed using a reference normal auditory model. To make the normal and impaired synthetic textures more comparable, parameters were transformed such that they varied linearly. The coefficient of variance was computed on the individual statistics. As can be seen in Fig. 3, the variation is not consistent for all textures suggesting that some statistical groups are more affected by changes in the early auditory processing than others.

Although it is valuable to compare the averaged variation in texture statistics between normal and impaired auditory models, it is perhaps more intuitive to examine the individual statistics for a given texture. Figure 4 shows this comparison for the sound texture *birds chirping*. The marginal statistics vary (Fig. 4A), particularly for the high frequency channels and higher-order marginal moments. However, for this texture, the time-averaged frequency spectrum is well preserved, as shown by the similarity between the normal and impaired mean statistics. The correlation statistics (Fig. figure:5B) vary as well, showing a noticeable increase in the co-variance of neighboring peripheral channels. This was expected for the hearing-impaired filters, as there is considerably more overlap between neighboring filters (see Fig. 2C). Lastly, the modulation power (Fig. 4C) reveals a difference between the two synthetic textures, particularly in the frequency region around 1.5 kHz for slow modulations.

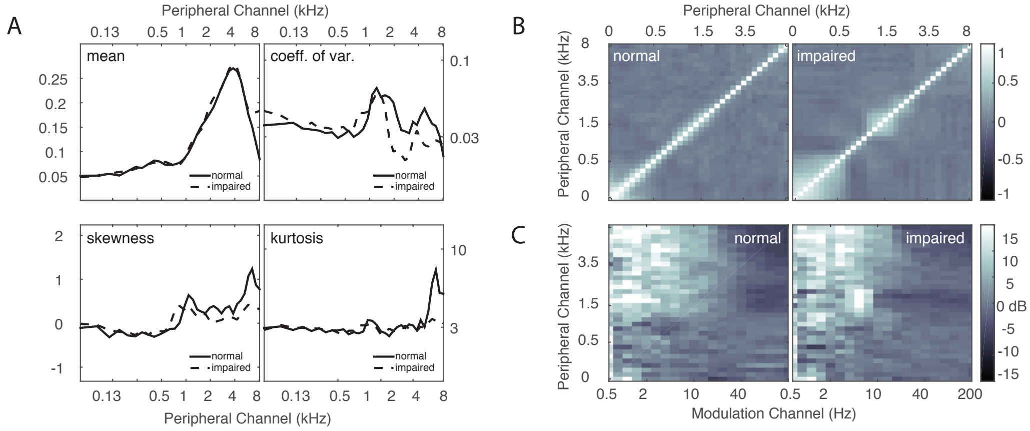

Fig. 4: Comparison of normal and imparied texture statistics for *birds chipring*. (A) Marginal moments (mean, coeff. of variance, skewness, kurtosis), (B) pair-wise correlations for subband envelope, (C) modulation power. Note the modulation pair-wise correlation statistics are not shown.

EXPERIMENTS

In order to investigate the significance of frequency selectivity and compression in sound texture perception, we asked listeners to discriminate between synthetic textures generated with normal and modified auditory models. The listeners were presented with three intervals, each 2 seconds in duration, and required to find the *odd* or modified interval, where two intervals were generated with a normal hearing model and the *odd* interval was generated with a modified hearing model. The stimuli were presented via open-ear headphones at a sound pressure level (SPL) of 65 dB. The modified texture could either be the first interval or the last interval. The two intervals generated from a normal hearing model were from the same texture family, but different sound instances, ensuring that listeners could not use unique acoustic features in their judgments.

Figure 5A shows the results for textures generated with broader as well as narrower peripheral filters, where the textures generated from ERB-spaced filters are the reference. Fifteen self-reported normal-hearing listeners participated in the experiment. The results show an increase in discrimination performance as the model deviated

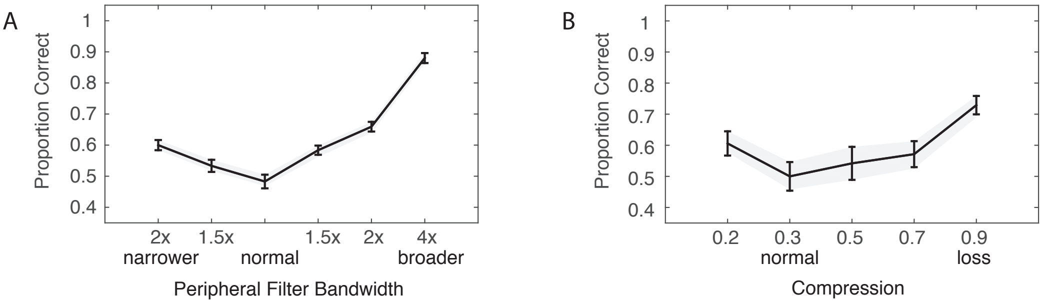

Fig. 5: Discrimination results for synthetic textures generated with impaired models of the auditory periphery shown as a proportion correct for (A) broader/narrower peripheral filter and (B) loss/change in compression. Error bars show standard error.

from the reference. This is particularly the case when the synthetic textures were generated with broader filters. However, it can also be seen that performance increases with narrower filters, suggesting that the higher number of filters may capture some additional frequency cues. Figure 5B shows the results for textures generated with reduced compression. Eight self-reported normal-hearing listeners participated in the experiment. The results show an increase in discrimination performance as the auditory model parameters deviated from normal hearing. The listeners reported audible artifacts in some of the intervals, and indeed, the change in compression seemed to offer cues when listening to modified compression settings. In addition, the synthesis process applies the compression during the analysis and removes the compression during the synthesis process, essential by reversing the effects of the compression. Therefore, the synthesis process seems to negate the possibility of exploring the perceptual consequences of compression with texture synthesis.

To better quantify the contribution of the texture statistics to the perception of normal and impaired synthetic textures, we designed a preference task experiment with stimuli that impaired particular statistical groups; marginal moments, pair-wise correlations, or modulation power. The listeners' were presented an original sound texture which was compared to two synthetic sounds generated from a normal and parametrically impaired auditory model. The three intervals were each 4 seconds in duration. The presentation of the synthetic intervals was randomized. The stimuli were presented via headphones at a level of 65 dB SPL.

The results from the parametrically impaired auditory model with 4x broader filters are shown in Fig. 6A. Twelve self-reported normal hearing listeners participated. The figure shows the pair-wise correlation parameter group was the most sensitive to impairment, as 72% of synthetic textures generated from a normal-hearing model were preferred over a pair-wise correlation-impaired model. The impaired marginal moments parameter group also showed an effect on the perception followed by the modulation power. It should be highlighted that a common modulation selective

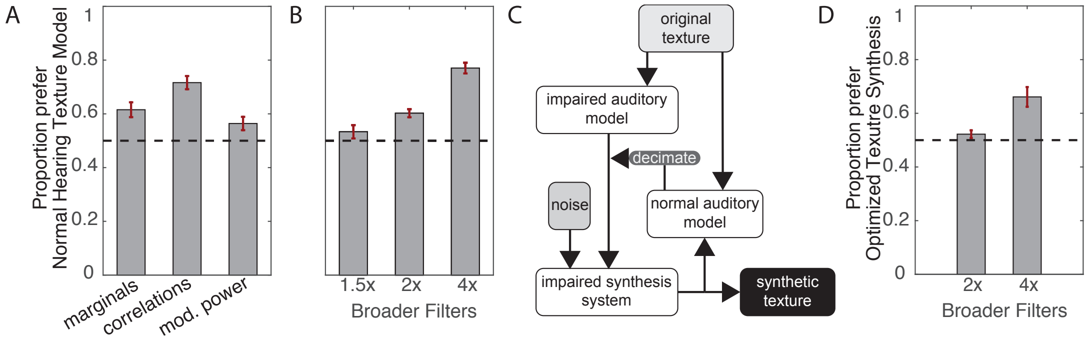

Fig. 6: Results of listeners who prefer textures synthesized with normal hearing model for (A) impairing individual parameter groups and (B) varying severity of the model impairment. (C) Compensation stategy for impaired sound texture synthesis. (D) Optimized impaired texture statistics results for 2× and 4× broader peripheral filters. Error bars show standard error.

filterbank structure was used for all synthetic textures. These results highlight the impact of the individual impaired parameter groups on hearing impairment.

As a control, we also asked listeners to perform a preference task with the wholly impaired auditory system with 3 configurations of peripheral filter broadening – 1.5×, 2×, and 4× – shown in Fig. 6B. The results are consists with the results shown in Fig. 5A as well as the parametrically varied impaired auditory model results. The results show that the perceptual quality declined as the auditory model deviated from that of a normal system.

COMPENSATION STRATEGY

Given that the representation and perception of synthetic sound textures change with the impairment of the peripheral auditory model, the question is whether it possible to modify the statistical representation to regain the perceptual quality towards the original texture. The results from experiments 1 and 2 revealed that a broadening of peripheral filters is salient for synthetic sound textures and most affected by the changes in the representation of pair-wise correlation statistics. A possible optimization strategy for an impaired auditory system could be a decimated version of the normal hearing statistics. However, the textures synthesized with an impaired model and decimated normal hearing statistics yielded poor synthetic versions, and often the synthesis failed. A different structure was implemented that used parallel normal and impaired model analysis systems, which is shown in Fig. 6C. The coupled analysis adjusts the impaired statistics such that the synthetic output is optimized to yield a synthetic texture similar to the original texture as measured by a normal auditory model. This can be thought of as *nudging* the impaired model representation to output a texture with similar perceptual qualities to the original texture.

Listeners performed a preference task to reveal the significance of the impaired auditory model optimization system. In each trial, listeners were presented with an original texture recording followed by two randomly presented synthetic textures; one synthesized with an impaired auditory model and another synthesized with the impaired auditory model optimization system. The stimuli were presented via headphones at a level of 65 dB SPL and each interval was 4 seconds in duration. The results from the impaired auditory model texture optimization system in Fig. 6D show a modest improvement in subjective performance for the 4× broader peripheral filter case. In the case of the 2× broader filters, no improvements were found. Although the performance of the optimization system did not yield comparable results to the original, there is modest benefit and the method does warrant further investigation.

SUMMARY

Sound textures offer a novel avenue for investigating the changes in representation due to hearing impairments, as well as the perceptual consequences of those changes. The differences in sound textures synthesized with auditory models that deviated from the normal hearing system were identifiable by normal-hearing listeners. The model impairments introduced changes to the statistical representation of sound textures, which related to perception to varying degrees. The results showed that pair-wise correlation statistics offer a primary auditory cue that affects the quality of the texture synthesis. Understanding how such *noise* signals are represented in the normal and impaired auditory system may offer some insight into the processing involved in "cocktail party" scenarios, where the auditory system separates a target signal from the noise.

REFERENCES

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier," J. Acoust. Soc. Am., **102**, 2892-2905.

Glasberg, B.R. and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res., **47**, 103-138.

Harte, J.M., Elliott, S.J., and Rice, H.J. (**2005**). "A comparison of various nonlinear models of cochlear compression," J. Acoust. Soc. Am., **117**, 3777-3786.

McDermott, J.H. and Simoncelli, E.P. (**2011**). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," Neuron, **71**, 926-940.

McDermott, J.H., Schemitsch, M., and Simoncelli, E.P. (**2013**). "Summary statistics in auditory perception," Nat. Neurosci., **16**, 493-498.

Moore, B.C.J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues.* John Wiley and Sons.

Rosengard, P.S., Oxenham, A.J., and Braida, L.D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am., **117**, 3028-3041.