EFFECT OF AUGMENTED AUDIFICATION ON PERCEPTION OF HIGHER STATISTICAL MOMENTS IN NOISE

Katharina Vogt, Matthias Frank, Robert Höldrich

Institute of Electronic Music and Acoustics,
University of Music and Performing Arts
Graz, Austria
{vogt, frank, hoeldrich}@iem.at

ABSTRACT

Augmented audification has recently been introduced as a method that blends between audification and an auditory graph. Advantages of both standard methods of sonification are preserved. The effectivity of the method is shown in this paper by the example of random time series. Just noticeable kurtosis differences are effected positively by the new method as compared to pure audification. Furthermore, skewness can be made audible.

1. INTRODUCTION

Sonification is still a relatively young field building up a canon of methodologies. Two of the standard approaches to sonification are audification and auditory graphs. Recently, the method of augmented audification has been introduced [1] as a seamless interpolation between these approaches. Augmented audification combines well-known techniques of signal processing: single-side-band modulation utilizing the Hilbert transform, and an exponential frequency modulation. This method allows to control both the mean position in the frequency range *and* the bandwidth of the sonification by free model parameters, independently to the rate of the data display. Fundamental properties of audification are conserved, notably the compact temporal support and the translation of high frequency content of the data into transient events in the sound. Furthermore, data sets can be explored interactively at various time scales and in different frequency ranges.

Frauenberger et al. [2] studied the audification of random data time series with varying higher order momentums. The third moment, skewness, is a measure for the asymmetry of the probability density function. The fourth moment is called kurtosis and serves as a measure of the peakedness of the distribution. In the study it has been shown that participants could discriminate a kurtosis difference in the audification of above 5. Qualitatively, they reported an increase of roughness with rising kurtosis. Distinguishing different values of skewness could not be proven. This is not surprising, as skewness depends strongly on the mean of the data series which results in an indiscernible DC value, and furthermore, human hearing does not perceive the "sign" of a signal.

In the following section, basic properties and limits of audification and auditory graphs are reviewed shortly. Sec. 3 introduces the signal processing algorithms of augmented audification[1]. Sec. 4 discusses the use case of random time series and the generation of data sets used in the listening experiment, which is discussed in Sec. 5. Finally, we conclude and give an outlook to further research.

[1] Accompanying sound examples can be found at: http://iaem.at/Members/vogt/augmentedaudification

2. COMPARISON OF AUDIFICATION AND AUDITORY GRAPHS

Audification, on the one hand side, has been defined by Kramer in 1991 (cited in [3], p. 186): "a direct translation of a data waveform to the audible domain". Today, audification has many "puristic" supporters within the sonification community who are in favor of a direct playback of data, with the only adjustable factor being the playback rate (see [4]). This factor also determines the typical size of data sets that are apt for audification, thus they are reasonably large.

A crucial advantage of audification is the following: by conserving the time regime of the data signal, audifications of real physical processes are usually broad-band with a pronounced proportion of high frequencies during rapid transients. In the task of identifying natural sounds, e.g., the attack of musical instruments or speech signals, the transient parts of the signal provide features for human hearing that serve as the basis for pattern recognition tasks. The same is true for audified signals.

In general, audification suffers from a trade-off between the macroscopic time scale and the frequency range of the relevant information. The ideal audification signal has relevant auditory gestalts within time and frequency regimes that can be well-perceived by the human auditory system. Some suggestions have been made to cope with this trade-off and argue in favor of a more adjustable audification paradigm (e.g., [5]).

On the other hand side, *auditory graphs* mostly sonify reasonably small data sets, up to a few hundred data points, often in a pitch-time-display. Thus, with respect to data size, we may find typical data of auditory graphs on the very other end of the scale than audification (e.g., on opposite sides on the sonic design space map, [6]). Obvious benefits of auditory graphs are the straightforward analogy to visual graphs, which make them intuitively understandable, at least for sighted users. The Sonification Sandbox [7] was possibly the largest effort to develop a general tool for auditory graphs. From the experience with the toolbox it can be concluded that most real-world sonification applications need a more flexible adjustment between the data set and the auditory graph.

Flowers [8] discussed promises and pitfalls of auditory graphs. He suggested that successful displays follow these strategies: numeric values should be pitch-encoded; the temporal resolution of human audition shall be exploited; loudness changes in a pitch mapped stream shall be manipulated in order to provide contextual cues and signal-critical events; distinct timbres shall be chosen in order to minimize stream confusions and unwanted perceptual grouping; and time in sound shall be used to represent time in the data.

Augmented audification allows to blend between audification and auditory graphs (in the form of a pitch-time display). Advantages of both methods can be combined, overcoming the time-scale trade-off of pure audification.

3. AUGMENTED AUDIFICATION: THE MODEL

For explaining Augmented Audification (henceforth: AugAudif), we start with a basic audification. We assume a data set $x(n)$ with $n = 1..N$ data points and a playback rate or sampling frequency f_p, i.e., f_p data points are displayed per second. The rendering over a D/A converter with a reconstruction filter leads to a continuous signal $x(t)$ with a bandwidth B between zero and $1/2f_p$ Hz. If the playback rate is as low as a few hundred data points per second, the resulting sound will be in a low frequency range, where the human ear is not very sensitive.

3.1. Frequency Shifting

Therefore, as a first step, we perform frequency shifting by a single-side-band modulation. Using a Hilbert transform (see, e.g., [9]), the original audification signal $x(t)$ becomes the complex-valued signal $x_a(t)$,

$$x_a(t) = x(t) + j\,\mathcal{H}\{(x(t)\} \tag{1}$$

with the imaginary constant j. This analytical signal can be written using a real-valued envelope $env(t) = |x_a(t)|$ modulated by a phasor with the instantaneous phase $\theta(t) = angle[x_a(t)]$:

$$x_a(t) = env(t)\, e^{j\theta(t)}. \tag{2}$$

Performing a frequency shift by Δf and taking the real part of this signal leads to a SSB-modulated sound signal $x_{SSB}(t)$:

$$\begin{aligned} x_{SSB}(t) &= Re\left[env(t)\, e^{j(\theta(t)+2\pi\Delta f t)}\right] & (3)\\ &= x(t)\cos(2\pi\Delta f t) - \mathcal{H}\{x(t)\}\sin(2\pi\Delta f t). & (4) \end{aligned}$$

The spectrum of the analytical signal, which contains (only non-negative) frequencies between zero and B Hz, is shifted to the range between Δf and $(\Delta f + B)$. Discarding the imaginary part re-builds a symmetric spectrum.

The frequency shift Δf is a free parameter of the method, which helps to yield a perceptually optimal frequency range of the sonification, i.e., somewhere within the range of 100 Hz and 2 kHz. If $\Delta f = 0$, there is no difference to a pure audification.

In the case of high playback rates, e.g., $f_p = 20$ kHz, which lead to a broad-banded audification, a frequency shift of $\Delta f = 100$ Hz hardly changes the overall signal, but might make low frequency components of the signal audible, as the spectrum is now shifted to the range between 100 Hz and 10.1 kHz.

A strong frequency shift, especially in combination with slow playback rates, results in a very narrow-banded signal which might be problematic from a perceptual point of view. The frequency shift squeezes the original - conceptually infinite - pitch range to a range of $(\Delta f + B)/\Delta f$. For example, if $f_p = 200$ Hz, hence the bandwidth of the primary audification signal is max. 100 Hz, and the spectrum is shifted by $\Delta f = 500$ Hz, the resulting bandwidth is 500 to 600 Hz. Speaking in musical terms, all frequency components of the original data stream are now concentrated within a minor third. Fluctuations of such narrow-banded signals are difficult to perceive.

3.2. Exponential Frequency Modulation

Therefore the method is extended by modulating the frequency of the phasor of the analytic signal $x_a(t)$. The instantaneous frequency of the modulator, $f_i(t)$, exponentially encodes the numeric data values of $x(t)$ as pitch, following to Flowers' recommendations:

$$f_i(t) = 2^{cx(t)} f_0. \tag{5}$$

f_0 is the carrier frequency and c a freely choosable parameter that controls the magnitude of the modulation: Setting $c = 0$ results in a constant instantaneous frequency of the frequency modulation which is then independent of the data values $x(t)$. This results in a pure frequency shift as described in Sec. 3.1. Setting $c = 1$ leads to a transposition of one octave higher/ lower for signal values $x(t) = +1/-1$. The value of c has to be chosen carefully depending of signal amplitude and bandwidth to prevent aliasing resulting from strong FM sidebands.

For the AugAudif, the parameter of frequency shift is used as carrier frequency, $f_0 \equiv \Delta f$. Integrating over the instantaneous frequency results in the instantaneous phase $\phi_i(t)$, which serves as a phase modulating term for the analytical signal.

$$\phi_i(t) = \int_0^t 2\pi\, \Delta f\, 2^{cx(\tau)} d\tau. \tag{6}$$

This leads to the complete model of Augmented Audification:

$$\begin{aligned} x_{AA}(t) &= Re\left[env(t)\, e^{j\,(\theta(t)+\phi_i(t))}\right] & (7)\\ &= x(t)\cos(\phi_i(t)) - \mathcal{H}\{x(t)\}\sin(\phi_i(t)). & (8) \end{aligned}$$

The model is controlled by two freely choosable model parameters, Δf and c, that can be set according to the explorative goals of the sonification.

4. USE CASE: STATISTICAL PROPERTIES OF RANDOM DATA TIME SERIES

4.1. Data generation

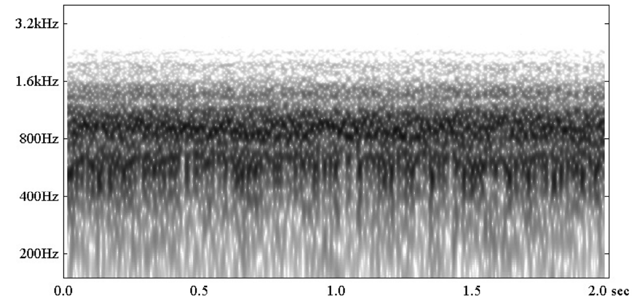
As a challenging use case for augmented audification, we generated random time series with different statistical properties. Frauenberger et al. [2] used a Levy alpha-stable distribution to create their data sets. The Levy alpha-stable distribution is not defined for a distribution parameter (skewness) of < 2 and for kurtosis values that are smaller than the Gauss distribution. Therefore, we implemented a type IV Pearson distribution for the generation of the examples in this section and the formal listening experiment in Sec. 5.

Neglecting a normalization constant, the type IV Pearson distribution is given as

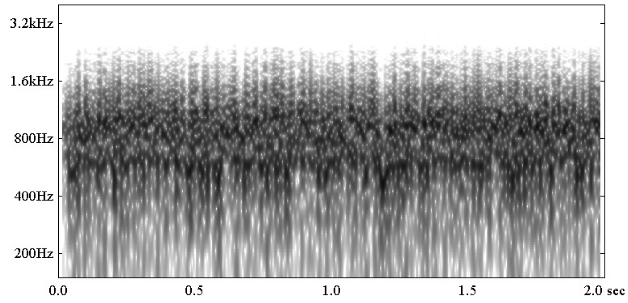
$$p(x) = \left[1 + \left(\frac{x-a}{b}\right)^2\right]^{-m} * e^{-\nu * \arctan\left(\frac{x-a}{b}\right)} \tag{9}$$

where a is the location parameter (mean value), b the scale parameter (variation), and the shape parameters m (kurtosis) and ν (skewness).

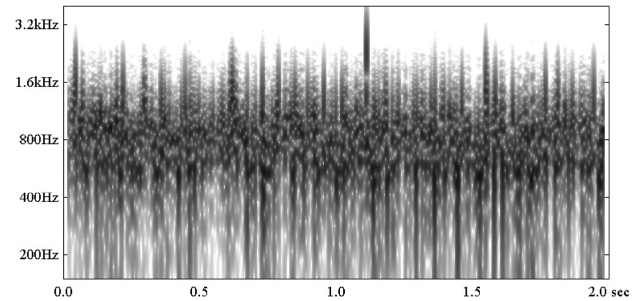
Noise samples of this distribution have been generated in Matlab for a mean value of zero and given standard deviation ($= 0.1$), but with various values for skewness and kurtosis. The distributions have been generated at a low sampling rate, thus limited in bandwidth to a cutoff frequency f_c.



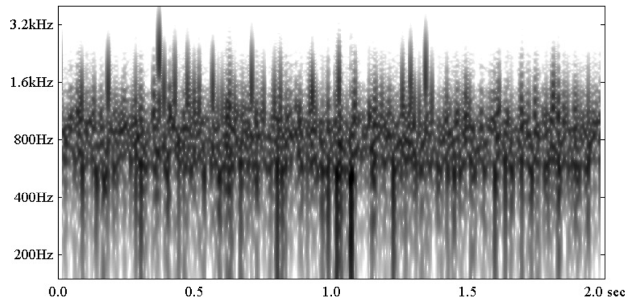
(a) Kurtosis = 1.



(b) Kurtosis = 2.



(c) Kurtosis = 4.



(d) Kurtosis = 8.

Figure 1: *Spectrograms of the 4 consecutive sounds in Soundfile 1. The fixed parameters are: skewness = 0;* $f_p = 800$ *Hz;* $c = 5/12$*;* $\Delta f = 600$ *Hz.*

Then, they have been interpolated to obtain the standard sampling rate of 44.1 kHz. Due to the random character of the generation process itself and the effects caused by the subsequent interpolation scheme, the actual statistical moments of the noise samples deviated from the target values. Therefore, an iterative procedure has been adopted to select noise samples with the intended preset values for skewness and kurtosis.

The actual values for skewness and kurtosis were calculated from the zero-mean samples $x(i)$, $i = 1..N$:

$$skew(x) = \frac{\mu_3}{\sigma^3} = \frac{\frac{1}{N}\sum_{i=1}^{N} x(i)^3}{\left(\frac{1}{N}\sum_{i=1}^{N} x(i)^2\right)^{3/2}}, \tag{10}$$

$$kurt(x) = \frac{\mu_4}{\sigma^4} = \frac{\frac{1}{N}\sum_{i=1}^{N} x(i)^4}{\left(\frac{1}{N}\sum_{i=1}^{N} x(i)^2\right)^{2}}. \tag{11}$$

4.2. Sound examples

A first, informal listening of the authors of this paper showed a much lower threshold for discriminating kurtosis and even the ability to defer different values of skewness using AugAudif as compared to direct audification. Two sound examples shall illustrate the effect of the method:

Soundfile 1 is an AugAudif of time series with a white noise spectrum, zero skewness and varying kurtosis (consecutively 1, 2, 4, and 8). The playback rate of the data has been chosen as 800 Hz. Fig. 1 shows the spectrograms of these 4 sounds. (All spectrograms were calculated using a 4096 sample Hanning window and are displayed on a logarithmic frequency scale up to 10 kHz.)

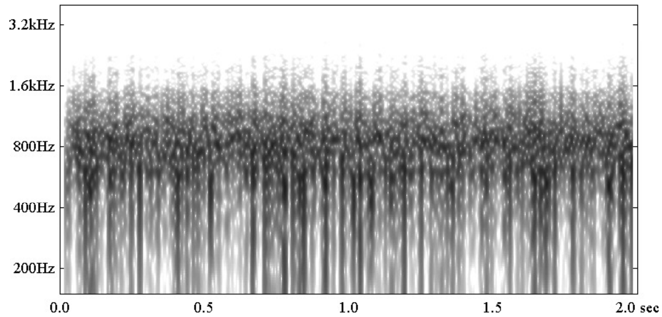
Soundfile 2 is an AugAudif of time series with constant kurtosis but varying skewness. The parameters are the same as above ($f_p = 800$ Hz; $c = 5/12$; $\Delta f = 600$ Hz). Kurtosis is set at 12, while skewness takes the values of -2, 0, and 2, respectively. The spectrograms shown in Fig. 2 clearly indicate the asymmetric frequency excursions due to the different skewness.

5. EXPERIMENT

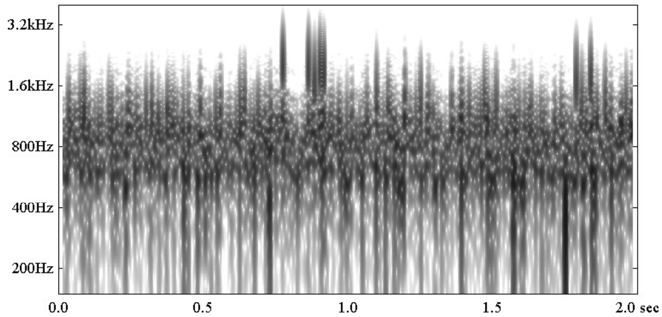
The above method has only recently been introduced. Evidence for its usefulness have to be found on an inter-subjective level. Therefore, a formal listening experiment was performed to test the following hypotheses:

- The just noticeable differences (JNDs) of kurtosis of random data time series may be lowered when treated with augmented audification as compared to direct audification.
- The skewness of random data time series may be detected when treated with augmented audification as compared to direct audification.

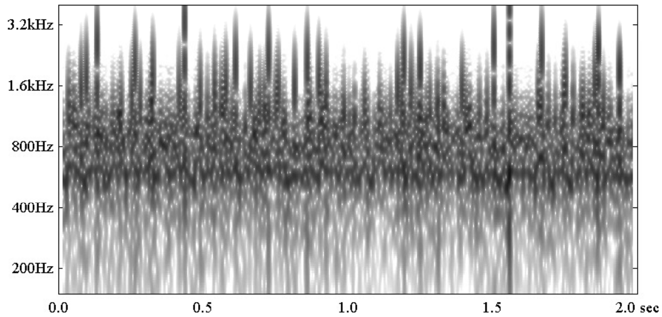
Additionally, the experiment should investigate the influence of different parameters for augmented audification on JNDs of kurtosis, i.e. provide estimates for optimal parameter settings in different frequency regimes.



(a) Skewness = -2.



(b) Skewness = 0.



(c) Skewness = 2.

Figure 2: *Spectrograms of the 3 consecutive sounds in Soundfile 2. The fixed parameters are: kurtosis = 12;* $f_p = 800$ *Hz;* $c = 5/12$*;* $\Delta f = 600$ *Hz.*

5.1. Setup, Method, and Conditions

The experiment evaluated the JNDs in skewness and kurtosis for 19 conditions shown in Table 1 using an adaptive 1-up/2-down triangle test [10]. Thus, the participant's task was to identify the odd one within a triplet of sounds. As all sound files had to be rendered before the experiment, the skewness/ kurtosis values have been quantized in discrete steps. For all parameters (conditions, quantized skewness/ kurtosis values) 5 different realizations of random time series have been created. During the experiment, the realizations have been selected in order that no triplet contained identical sound files (but two realizations with the same parameters, and one other). In this way, the experiment considered the perception of variation within realizations with same parameters as well. Participants could listen to the sounds as often and in any order they wished.

As a reference value for all conditions with varying kurtosis, a kurtosis of 3 was taken, which corresponds to Gaussian noise; within mathematics it is common to use this reference (with the notions of platykurtic or leptokurtic distributions referring to a negative or positive excess related to a kurtosis of 3). Furthermore it seems plausible from an evolutionary point of view that the auditory system is "gauged" to the normal distribution as a frequent case in natural systems.

In order to minimize the number of iterations in the adaptive procedure, a variable step size has been used: each procedure started with a comparison of a parameter set with largest available skewness/ kurtosis value and one with the reference value. The step size (difference of skewness/ kurtosis value to reference value) was halved after each two correct answers in a row and after each wrong answer. The new step size was calculated from the old one, plus/ minus half of the actual step size. In case of all correct answers, and due to the quantization of the skewness/ kurtosis values, the smallest step size could be achieved after 7 iterations. The maximum number of iterations has been limited in dependence of the number of quantization steps, cf. Table 1. This measure should prevent the participant from fatigue and account for cases where s/he could not discriminate between different settings at all.

In addition to the psychometric measurement of the JNDs, participants were asked to verbally describe the perceived differences in an accompanying questionnaire. For each condition, the participants should do so right after the first triplet, when the differences were as large as possible.

For playback of the sound files, an RME Multiface and Beyerdynamic DT-770 Pro headphones have been employed. A total of 10 subjects (all experienced listeners with hearing loss of less than 15dB, 8 of them part of a trained expert listening panel [11, 12, 13]) participated in the experiment. On average, the whole experiment took 115 min. and was divided into two parts.

Table 1: *Conditions in the experiment with variable or fixed values for skewness and kurtosis (reference values in brackets), number of quantization steps, and maximum number of iterations.*

con.	f_c/Hz	Δf/Hz	c	kurtosis	skew.	steps	iter.
1	10000	-	-	var(3)	0	12	25
2	5000	-	-	var(3)	0	12	25
3	500	-	-	var(3)	0	12	25
4	100	-	-	var(3)	0	12	13
5	5000	150	5/12	var(3)	0	12	25
6	500	150	5/12	var(3)	0	12	25
7	100	150	5/12	var(3)	0	12	25
8	5000	600	5/12	var(3)	0	12	25
9	500	600	5/12	var(3)	0	12	25
10	100	600	5/12	var(3)	0	12	25
11	5000	600	3/12	var(3)	0	12	25
12	500	600	3/12	var(3)	0	12	25
13	100	600	3/12	var(3)	0	12	25
14	5000	-	-	var(1.27)	0	16	25
15	5000	-	-	var(2.1)	0	14	25
16	5000	-	-	var(3.5)	0	10	20
17	5000	-	-	var(4.5)	0	8	18
18	500	600	5/12	8	var(0)	5	13
19	5000	600	5/12	8	var(0)	5	13

5.2. Quantitative Results

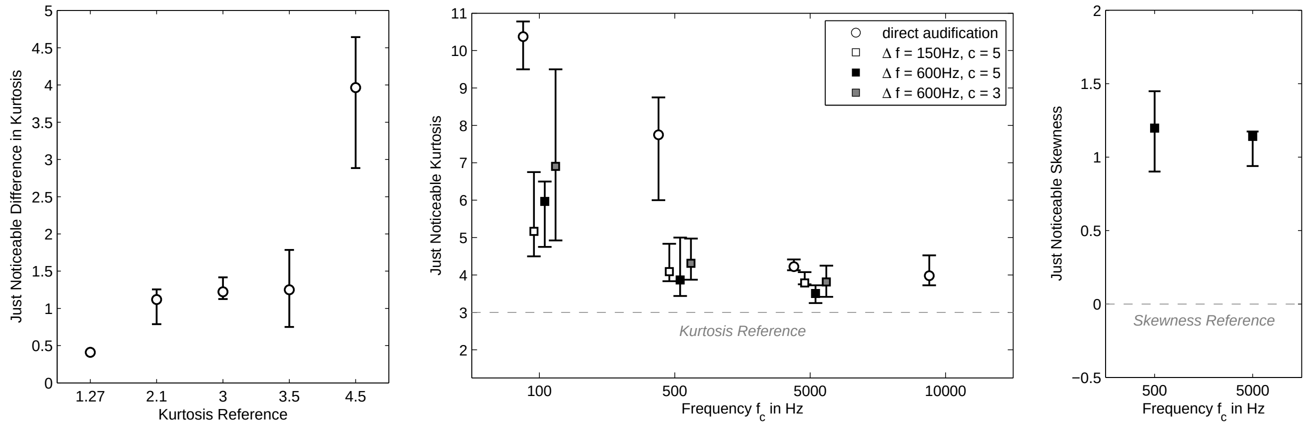
The above procedure lead to a standard oscillating between two values of smallest step size that supposedly lie around the JND. The thresholds for the just noticeable values were calculated as the minimum of the averages over the last 4 and 6 reversals [14]. For condition 4, this calculation did not always converge. Some participants never answered two times correctly in straight succession. In this case, the just noticeable value has been set to the maximum available value for this condition. Similarly, most participants could always perceive the smallest available difference in condition 14. Thus, the just noticeable kurtosis difference was set to the smallest available difference. In comparison to our results, the actual just noticeable difference might hence be larger for condition 4 and smaller for condition 14. However, this would not change the general statements of our study.

Figure 3a shows the different just noticeable differences in kurtosis for direct audification in dependence of the reference kurtosis. Analysis of variance and Kruskal-Wallis tests revealed a significant increase towards higher reference kurtosis values ($p < 0.001$). This also holds for the variance of the just noticeable difference. However, only the neighboring conditions 14/15 and 16/17 yield significantly different means ($p < 0.001$) and median values ($p \leq 0.002$). The results indicate varying JNDs for different values of reference kurtosis. However, a modeling of the difference threshold departing from Weber's law could not be deduced from the data. Nevertheless, a first attempt to establish a psychophysical scale for kurtosis of band-limited noise in direct audification can be found in the Appendix.

Figure 3b encompasses the central findings of the experiment. Just noticeable kurtosis decreases for higher f_c using either direct or augmented audification. For all evaluated cutoff frequencies f_c, AugAudif yields smaller just noticeable kurtosis as compared to direct audification. The different variations of augmented audification perform similarly well, although $c = 3/12$ tends to be the worst variation for $f_c = 100$ Hz and $\Delta f = 600$ Hz, $c = 5/12$ tends to be the best variation for $f_c = 5000$ Hz. The statistical analysis revealed that kurtosis sensitivity increases significantly for higher f_c ($p < 0.001$) in direct audification.
All neighboring conditions yield significantly different mean ($p < 0.001$) and median values ($p \leq 0.001$), except for conditions 2/1 ($p \geq 0.54$). Just noticeable kurtosis also decreases for AugAudif ($p \leq 0.04$), although conditions 6/5 yield no significantly different mean ($p = 0.15$) and median values ($p = 0.18$). It can be argued, that the magnitude of the difference between augmented and direct audification depends on f_c in relation to Δf: while $\Delta f = 150$ Hz is a large difference compared to $f_c = 500$Hz, it is relatively small compared to $f_c = 5000$ Hz. As might be expected, the conditions where $c = 3/12$ tend to lead to worse results than the ones with $c = 5/12$. This can be argued from the sensitivity of the human hearing, which, in general, decreases for lower frequencies. For $f_c = 100$ Hz, all variations of augmented audification yield significantly lower median values ($p \leq 0.002$) compared to direct audification. However, comparing the different variations among each other, none of them show significant differences in means ($p \geq 0.11$) and median values ($p \geq 0.1$). Condition 13 suffers from a remarkably strong inter-subjective variation.

The authors would have expected that direct audification saturates from the low frequency range towards the most sensitive hearing range (even if the value for high frequencies, i.e. the 10 kHz regime, should exhibit a degradation following this argument, which is not reflected in the data). Similarly, for $f_c = 500$ Hz, all variations of augmented audification yield significantly lower median values ($p \leq 0.001$) in comparison to direct audification. Again, there are no significant differences between the means ($p \geq 0.31$) or median values ($p \geq 0.29$) of the different variations. For $f_c = 5000$ Hz, although all augmented audifications yield significantly smaller means ($p \leq 0.07$) and median values ($p \leq 0.07$), condition 5 achieves significantly smaller means/medians ($p = 0.02$) than condition 8. However, the difference between conditions 8 and 11 is not significant ($p = 0.11$). The optimal parameter setting within this experiment has been reached for condition 8, with $f_c = 5000$ Hz, $\Delta f = 600$ Hz, and $c = 5/12$. In percentages, this corresponds to a frequency shift in the order of 8% of the mean frequency range, and a factor c equalling roughly one percent of the frequency shift.

Considering pure audification of band-limited noise signals, the sensitivity for kurtosis depends strongly on the bandwidth. The increase of sensitivity up to a cutoff frequency of a few kHz can be well-explained by general sensitivity properties of the human auditory system. The kurtosis sensitivity shows a saturation effect above 5 kHz which can be attributed to the fact that high kurtosis values result in a very peaked signal and therefore in a broadband "synchronization" of energy portions up to very high frequencies, but the further information contained in the extended bandwidth cannot be exploited by the auditory system due to masking and the reduced sensitivity above 5 kHz. The JND of kurtosis also depends on the kurtosis reference. Nevertheless, our attempts to establish a Weber's model for the JND dependence on the stimulus' value failed for various reasons. First, such models require an absolute zero value for the stimulus, but kurtosis lacks a natural and/or plausible zero point although there is a mathematically defined minimum value of 1. (A kurtosis of 1 could be achieved because of the band-limited nature of the noise signals.) Secondly, the perceptive attribute correlated with varying kurtosis is not a prothetic continuum like sound intensity or brightness, as is also supported by the qualitative analysis.



(a) direct audification at cutoff frequency $f_c =$ 5000 Hz (conditions 14,15,2,16,17)

(b) direct and augmented audification with different Δf, c (conditions 4,7,10,13, 3,6,9,12, 2,5,8,11, 1)

(c) augmented audification with $\Delta f = 600$ Hz, $c = 5$ (conditions 18,19)

Figure 3: *Median and corresponding 95% confidence intervals of just noticeable skewness/ kurtosis with regard to reference skewness/ kurtosis.*

The results in Figure 3c show that variation of skewness is audible. However, f_c has no significant effect on just noticeable skewness variation ($p = 0.59$ for means, $p = 0.49$ for median values). The reference case for direct audification has not been studied in this experiment, but in informal listening of the authors prior to the experiment and by Frauenberger et al. [2]. For detecting skewness, the pure frequency shift should not lead to an improvement, whereas by the additional frequency modulation the asymmetric behavior of the distribution values becomes apparent.

5.3. Qualitative Results

The accompanying questionnaire was analyzed following [15] with regard to possible different criteria that the subjects may have used in different conditions. Furthermore the analysis should help defining common terms for perceptive attributes related to skewness and kurtosis.

In general, a very diverse set of terms have been mentioned: only two categories have been subsumed in the first place, these are terms related to the (in-)homogeneity of the noise and its frequency bandwidth. The category of (in-)homogeneity included, among others, the notions ”static sound”, ”variation”, ”fluctuation”, ”outliers”, or ”(un-)evenness”. Notions related to frequency bandwidth were ”high/ low frequency”, ”f-spread” etc. Apart from these two groups of terms, subjects indicated 47 different adjectives describing sound attributes, plus another 19 noun groups indicating sound sources from a natural or technical context. These terms could not be further grouped: they consist of a quite extensive list of notions describing different noise-like sounds, often using colloquial words[2].

The terms were annotated (by numbers 1 to 47, the natural sound sources as N1 to N19) and grouped according to different experiments’ parameters. Apart from the two term groups above, the top five were ”knistern” (crinkle/ crackle), ”britzeln” (colloquial term, something like crumbling), ”blubbern” (bubbling), ”körnig” (granular), and ”knacksen” (crackle/ click). The most common sound sources cited are related to water drops and different types of rain, but a few participants were very creative here (e.g., *“aggressive electro-mosquitos”* or *”futuristic laser-bursting bubbles”*).

Cross-subject coincidences could be found in 11 out of the 18 conditions (defined by 4 or more participants using the same word/ category). The group category of (in-)homogeneity (4 times) or frequency bandwidth (one time) was used commonly the most often. ”Knistern” (crinkle/ crackle) was the only other frequently shared term (used 4 times), but all in the 5000 Hz conditions, thus the term refers only to higher frequencies and cannot be generally applied; ”blubbern” (bubbling) and ”knacksen” (crackle/ click) were commonly used in one experiment each (conditions 5 and 8).

In general about half of the subjects use rather uniform vocabulary, whereas the other half indicated a wide spread of diverse terms. Therefore the applied method of counting common words has to be treated with caution. Still, a few findings are interesting. For pure audification, we find the most intersubjectively repeated categories, i.e., it can be argued that the augmented audification results in more varied sounds. At low frequency ($f_c = 100$Hz) the general category of (in-)homogeneity is by far most important (very common for 4 subjects). It can be argued that it is harder to differentiate sound properties in the low frequency range, where we are less sensitive. On the other hand, the higher the frequency the more varied the terms become, and the analysis will neither lead to coincidences of commonly used terms.

The factor c seems to not have a large influence, at least comparing $c = 3/12$ to $c = 5/12$. The indicated terms stayed the same for more than half of the subjects within the conditions with the same Δf and f_c but varying c. As skewness was only varied in two experiment settings, no different wording could be found there.

Overall, the above analysis supports to introduce ”crackling” as the most general term to describe the influence of kurtosis in noise.

[2] Therefore the authors of this paper cannot provide a translated version of the original data in this paper; a specialized translator would be needed for this task. An example are the two German terms ”knittern” and ”knistern” (crinkle/ crackle), that are not only phonetically similar but have also a very similar meaning. Thus, it cannot be inferred that subject A has the same notion of ”knistern” as subject B has of ”knittern”.

6. CONCLUSIONS AND OUTLOOK

Augmented audification allows to interpolate seamlessly between pure audification and an auditory graph in the form of a pitch-time-display. As opposed to pure audification, where only the playback rate can be changed, two more model parameters can be chosen independently in the augmented audification. One parameter, Δf, controls the magnitude of a frequency shift. The second, c, sets the excursion of the exponential frequency modulation (FM), i.e. pitch modulation.

The effectivity of augmented audification has been shown in this paper by a listening experiment. The two hypotheses given in Sec. 5 are supported by the results of this experiment. When treated with augmented audification as compared to direct audification, in random data time series both the JNDs of kurtosis may be lowered and skewness may be detected. Furthermore, qualitative data on the denotation of a perceptive attribute related to skewness and kurtosis have been collected and analyzed. The English word "crackling" seems to be a promising candidate for intersubjective naming.

Cautiously, the optimal parameter settings for augmented audification can be given by a frequency shift in the order of 8% of the mean frequency range, and a factor of c equalling roughly one percent of this frequency shift. In general, the method of augmented audification has to be studied by other examples than random time series.

The regime between $f_c = 500$ and 5000 Hz is most interesting for further research, in order to get more information on the saturation of the JNDs with rising frequency and how the perceptual function might be modeled. Aspects of interactivity in the research paradigm (e.g., the possibility to acoustically zoom into an audified data set) have to be examined further.

7. ACKNOWLEDGMENT

We would like to thank all participants for their sustained effort in the (noisy!) listening experiment.

8. REFERENCES

[1] Robert Höldrich and Katharina Vogt, "Augmented audification," in *Proc. of the International Conference on Auditory Display (ICAD)*, 2015.

[2] Christopher Frauenberger, Alberto deCampo, and Gerhard Eckel, "Analysing time series data," in *Proc. of the International Conference on Auditory Display (ICAD)*, 2007.

[3] Gregory Kramer, Ed., *Auditory Display. Sonification, Audification and Auditory Interfaces.*, Proceedings Volume XVIII. Santa Fe Institute, Studies in the Sciences of Complexity, 1994.

[4] Florian Dombois and Gerhard Eckel, "Audification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J.G. Neuhoff, Eds., chapter 12, pp. 237–272. Logos Publishing House, Berlin, 2011.

[5] David Worrall, *Sonification and Information – Concepts, Instruments and Techniques*, Ph.D. thesis, University of Canberra, 2009.

[6] Alberto de Campo, "Toward a data sonification design space map," in *Proc. of the International Conference on Auditory Display (ICAD)*, 2007.

[7] Benjamin K. Davison and Bruce N. Walker, "Sonification sandbox reconstruction: Software standard for auditory graphs," in *Proc. of the International Conference on Auditory Display (ICAD)*, 2007.

[8] John H. Flowers, "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions," *Proceedings of International Conference on Auditory Display*, 2005.

[9] Alan Oppenheim and Ronald Schafer, *Digital Signal Processing*, Prentice Hall, 1975.

[10] Harry Levitt, "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 467–477, 1971.

[11] Alois Sontacchi, Hannes Pomberger, and Robert Höldrich, "Recruiting and evaluation process of an expert listening panel," in *Fortschritte der Akustik, NAG/DAGA*, Rotterdam, 2009.

[12] Matthias Frank, Alois Sontacchi, and Robert Höldrich, "Training and guidance tool for listening panels," in *Fortschritte der Akustik, DAGA*, Berlin, 2010.

[13] Matthias Frank and Alois Sontacchi, "Performance review of an expert listening panel," in *Fortschritte der Akustik, DAGA*, Darmstadt, 2012.

[14] Sygal Amitay, Amy Irwin, David J. C. Hawkey, Justin A. Cowan, and David R. Moore, "A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listenersa)," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1616–1625, 2006.

[15] Philipp Mayring, *Qualitative Inhaltsanalyse*, Beltz Verlag, 11. edition, 2010.

[16] GÖSta Ekman, "Weber's law and related functions," *The Journal of Psychology*, vol. 47, no. 2, pp. 343–352, 1959.

[17] Kenneth H Norwich and Willy Wong, "Unification of psychophysical phenomena: The complete form of fechner's law," *Perception & Psychophysics*, vol. 59, no. 6, pp. 929–940, 1997.

[18] John Z Sun, Grace I Wang, Vivek K Goyal, and Lav R Varshney, "A framework for bayesian optimality of psychophysical laws," *Journal of Mathematical Psychology*, vol. 56, no. 6, pp. 495–501, 2012.

[19] David Worrall, "The use of sonic articulation in identifying correlation in capital market trading data," in *Proc. of the International Conference on Auditory Display (ICAD)*, 2009.

Appendix:

In this supplementary discussion, we investigate the construction of a psychophysical scale for kurtosis of band-limited noise for the case of 5 kHz bandwidth and direct audification (see Table 1, conditions 14, 15, 2, 16, and 17).

In general, a monotonic psychophysical scale of sensation Ψ as a smooth function of the stimulus ϕ can be established from measurements of stimulus difference thresholds $\Delta\phi$ if a model for the subjective JND $\Delta\Psi$ is assumed. The most important JND models are Fechner‘ s law of a constant subjective JND, $\Delta\Psi = 1$, and Ekman‘s principle [16] which states a constant relative JND, $\Delta\Psi/\Psi = k$.

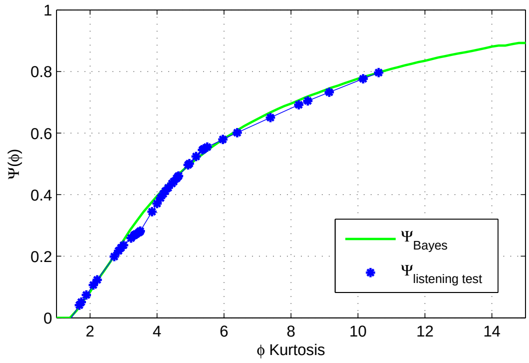


Figure 4: Psychophysical scale for kurtosis.

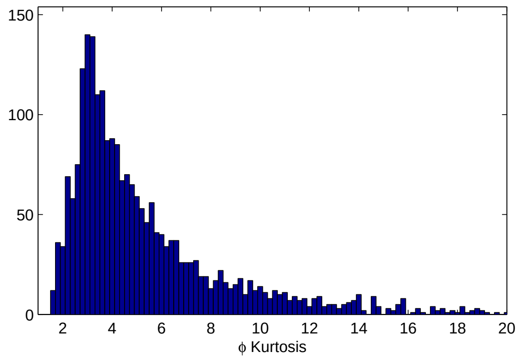


Figure 5: Kurtosis histogram of 20 min. of human speech.

However, though there has been a long-lasting debate about which of the two “laws” might be more appropriate, they are interrelated with the mathematical structure of the underlying psychophysical scale [17]. Considering an explicitly given or postulated psychophysical magnitude function $\Psi(\phi)$ and its derivative $d\Psi(\phi)/d\phi$, one can substitute finite differences for the differential

$$\frac{d\Psi(\phi)}{d\phi} = \frac{\Delta\Psi}{\Delta\phi}. \tag{12}$$

Assuming constant relative JND, $\Delta\Psi = k\Psi$, the psychophysical scale Ψ can theoretically be established utilizing numerical integration of the difference thresholds

$$\Psi(\phi) = exp\left[\int \frac{k * d\phi}{\Delta\phi}\right]. \tag{13}$$

Alternatively considering Fechner’s model, the scale function reads $\Psi(\phi) = \int(k * d\phi/\Delta\phi)$. Thus, difference threshold $\Delta\phi$, JND $\Delta\Psi$ and psychophysical scale Ψ are intertwined as functions of the stimulus ϕ.

The difference thresholds of our listening test have not been measured with enough density on the stimulus scale. Therefore, and for intersubjective differences, the calculation of $\Psi(\phi)$ has to be achieved by optimizing a cost function based on a certain assumption of JND model. We chose constant relative JND. Combing five reference values $\phi_j = 1.27, 2.1, 3, 3.5, 4.5$ with the difference thresholds of our ten subjects results in 42 different kurtosis values and therefore in 42 values of the psychophysical scale $\Psi_i = \Psi(\phi_i)$ to be modeled. (In some conditions, thresholds for different subjects yielded identical values.) For each reference value ϕ_j and for each of the associated individual difference thresholds $\phi_{j,m}$, Ekman‘s law postalutes: $\Psi_{j,m} = (1+k) * \Psi_j$ which enters the optimization problem as a quadratic cost function $J = \sum(\Psi_{j,m} - (1+k)\Psi_j)^2$.

Because of the different $\phi_{j,m}$-values for a single reference kurtosis, the resulting function $\Psi(\phi)$ would exhibit plateau regions and perhaps rather steep transitions between them which conflicts the required smoothness. Therefore, the cost function was extended by a smoothness term based on the distance in stimulus between adjacent scale values,

$$\lambda \sum_{i=1}^{40} \left(\frac{\phi_{i+2} - \phi_{i+1}}{\phi_{i+2} - \phi_i}\Psi_i - \Psi_{i+1} + \frac{\phi_{i+1} - \phi_i}{\phi_{i+2} - \phi_i}\Psi_{i+2}\right)^2,$$

and by a term setting a reference scale value as $\Psi(\phi_{ref}) = \Psi_{ref}$.

Optimization results have been obtained for different smoothness weights λ and a wide range of JND constants $k \approx 0.2 - 0.8$. They revealed minor sensitivity to the actual choice of parameter values. A typical scale function for $\lambda = 50$ and $k = 0.4$ is displayed in Fig. 4 (solid line with asterisks).

To test the psychophysical plausibility of our scale function, we compare it with the predictions of a theoretical model for limited perception proposed by Sun et al. [18]. Their approach relies on the neurophysiologically well-motivated assumption that the acquisition of information in a stimulus is Bayes-optimal at a computational level. Generally, it is argued that the actual mapping function $\Psi(\phi)$ is the outcome of an optimization process taking place during evolution. The optimization minimizes the expected relative error of the quantized representation of the stimulus and therefore depends on its probability density function. Assuming a stimulus pdf $f_{kurt}(\phi)$ and a bounded stimulus range $0 < \phi_0 \leq \phi \leq \phi_1 < \infty$, the psychophysical scale function has to satisfy

$$\frac{d\Psi(\phi)}{d\phi} \propto \phi^{-2/3} f_{kurt}(\phi)^{1/3} \quad for \quad \phi \in [\phi_0, \phi_1] \tag{14}$$

according to the proposed model (cf. Sun et al).

To compare Sun‘s model with our psychophysical scale function, a probability density function for kurtosis reasonably motivated from an ecologial and evolutionary point of view has to be established. Following Sun et al., we argue that the auditory system may have evolved to optimally process vocalization sounds like human speech. Therefore, we estimated $f_{kurt}(\phi)$ for human speech based on frames of 100ms length in 20 min of different speech recordings (the histogram is shown in Fig. 5) and calculated the Bayes-optimal psychophysical scale function $\Psi_{Bayes}(\phi)$. A direct comparison between the scale model based on the listening test and the Bayes-optimal model is displayed in Fig. 4. The values of our model have been normalized to match the range of the Bayes model. Though the two functions fit surprisingly well which at least indicates the appropriateness of this line of argumentation, it has to be stated that this is neither a direct proof for the validity of Ekman‘s principle of constant relative JND nor for the Bayesian perception model. In fact, it primarily reveals that the constant relative JND corresponds to the assumption of the relative error criterion in the Bayes model.