Focused Audification and the optimization of its parameters

Katharina Groß-Vogt1 • Matthias Frank1 • Robert Höldrich1

Received: 27 May 2019 / Accepted: 6 December 2019
© The Author(s) 2019

Abstract
We present a sonification method which we call Focused Audification (FA; previously: Augmented Audification) that allows to expand pure audification in a flexible way. It is based on a combination of single-sideband modulation and a pitch modulation of the original data stream. Based on two free parameters, the sonification’s frequency range is adjustable to the human hearing range and allows to interactively zoom into the data set at any scale. The parameters have been adjusted in a multimodal experiment on cardiac data by laypeople. Following from these results we suggest a procedure for parameter optimization to achieve an optimal listening range for any data set, adjusted to human speech.

Keywords
Sonification • Auditory display • Audification • Auditory graph • ECG data

1 Introduction
Sonification is still a young field building up a canon of methodologies, e.g. for supporting multimodal displays. Two of its standard methods are audification and auditory graphs. Audification is “a direct translation of a data waveform to the audible domain” (cited in [16], p. 186). It is often used to display one-dimensional, large data sets (with data display rates of tens of kHz).
2 Discussion of existing methods

2.1 Audification

Audification is one of the oldest methods of sonification research. A prominent, early study on the audification of seismic signals was conducted by Speeth et al. [23]. Subjects showed up to 90% discrimination rates between the sounds of earthquakes from the ones of atomic bombs.

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 signal portions provide important and salient features for the human ear and thus should serve as basis for pattern detection or recognition tasks in the auditory data exploration. Many authors, e.g., Dombois and Eckel [6], have argued in favor of a puristic approach to audification with as little data preprocessing as possible. This strategy should maximize the potential of the human hearing to detect yet unknown structures in the data which might be impaired by more sophisticated preprocessing.

The ideal audification signal has relevant auditory gestalts within time and frequency regimes that can be well-perceived by the human auditory system. Data sets that are problematic in audification shall be discussed with a thought experiment: let us assume a data stream with transient events that appear within a range of 1 k data points and with an (aperiodic) interval of roughly 10 k data points. This hypothetical data set is shown in Fig. 1. With a playback rate of 44.1 kHz (Fig. 1a), we find approximately four of these events per second, which is comparable to the rate of syllables per second in English spoken language² and thus apt for human hearing. On the other hand, each transient event lasts for approximately 22 ms and appears as a band-limited impulse with a cut-off frequency of around 50 Hz, which is far below the most sensitive frequency range of the hearing system. If the playback rate were to be raised by, e.g., a factor of 10, see Fig. 1b, the individual impulses would be transposed to a more appropriate frequency range, but at the cost of an indiscernible temporal structure of the impulse series. Concluding from this example, pure audification may suffer from a trade-off between the rhythmic structure and the displayed frequency range of individual events.

Different concepts have been elaborated to cope with this trade-off. Worrall [28] extends the notion of audification, and allows other means of data pre-processing: besides filtering and data interpolation, i.e. compression and frequency shifting. This wider definition of audification still excludes the explicit synthesis of sound or the use of specific signal models (such as the one we develop in this paper). Another similar approach has been explored within the CoRSAIRe project for a multisensory virtual reality environment [25]. In this project, “sonification–audification metaphors” have been developed for a specific data set (computational fluid dynamics, data), notably three different ways of shifting the pitch of a signals without changing its duration. Each of these algorithms has specific artifacts changing the audio outcome, but all have the advantage of adapting the audified data set better to the human listening range. This argument meets our motivation for FA, but we use a different algorithm (leading to different artifacts). Equally favourable to FA, Feedback from the data experts involved in the CoRSAIRe project found a fourth method based on FM more intuitive than the approaches using pitch shifting.

2.2 Auditory graphs

Just as audification, auditory graphs belong to the standard repertoire of sonification research since its beginning. Obvious benefits are the straightforward analogy to visual graphs, which make them intuitively understandable for sighted users, and their accessibility for non-sighted users. The data sets used are normally small, up to a few hundred data points, but may have several dimensions, as human audition is apt to segregate parallel data streams [11]. Focused Audification, depending on the setting of its parameters, may lead to a sonification that resembles in many ways an auditory graph (i.e., data points are mapped to pitch over time). Therefore, we shortly discuss the design of auditory graphs in the following.

Earlier research [3,15] recommended using musical pitches (e.g., MIDI notes following the Western 12-tone scale) mapped to the y-axis and time to the x-axis. The Sonification Sandbox [4] was possibly the largest effort to develop a general tool for auditory graphs in such a way. From experience

---

² This is a rule of thumb according to [27]. Vowels per second is a good estimation for syllables per second, which is correlated to the phoneme rate in normal rate speech. An automatic estimation of speaking rate (SR) for different languages measured the SR between 3.0 (for Mandarin) and 4.9 (for Japanese) vowels per second for different languages (English: 3.8) [19].
Fig. 1 Hypothetical data showing a possible trade-off between rhythmic structure and frequency in audification. The plots on top, show a time signal and power spectrum of a hypothetical data set containing some background noise and transient, low frequency events of approximately 22 ms each, with energy in the lower frequency domain well below 500 Hz. The lower plots, show an acceleration of the audification by a factor of 10. Now, the spectrogram shows a broad-band signal, but single events take place at very fast rates and are hardly discernible.

with the toolbox it can be concluded that most real-world sonification applications need a more flexible software environment. Also, we consider the limitation to 12 semi-tones and (aesthetically unpleasant) MIDI-sounds not any more state-of-the-art. A recent development of a general-purpose tool for sonification is the sonification workstation (see [20] also for a discussion of previous attempts).

In an analysis, Flowers [8] discussed promises and pitfalls of auditory graphs. He suggested the following strategies for successful displays:

- Pitch coding of numeric value
- Exploiting temporal resolution of human audition
- Manipulating loudness changes in a pitch mapped stream to provide contextual cues and signal-critical events
- Using time to represent time

All strategies but the last one are taken into account in the design of the proposed method: the last point, using time to represent time, might be fulfilled depending on the data set.3

Footnote 3 continued

3 Furthermore, in the case of several data sets, Flowers suggests to choose distinct timbres to minimize stream confusions and unwanted perceptual grouping and, in general, to compare sonified data sequentially rather than simultaneously.

3.1 Frequency shifting

Therefore, as a first step, we perform frequency shifting using a single-side-band (SSB) modulation. Applying a Hilbert transform \( \mathcal{H} \) (see, e.g., [18]), the original audification signal \( x(t) \) becomes the complex-valued signal \( x_a(t) \):

\( x_a(t) = x(t) + j\mathcal{H}(x(t)) \)

This transformation allows for easier manipulation of the audification data, enabling the separation of different frequency components.

3 Focused audification: the model

For explaining Focused Audification (henceforth: FA), we start with a simple audification. We assume a dataset \( x(n) \) with \( n = 1 \ldots N \) data points of a constant sampling frequency \( f_s \). In the most direct audification we take \( f_s \) equivalent to the playback rate \( f_p \), i.e., \( f_s \) 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 \( \frac{1}{2} f_s \) Hz. If \( f_s \) and therefore \( f_p \) is as low as a few hundred data points per second, the resulting audification will be in a low frequency range, where the human ear is not very sensitive.
\[ x_a(t) = x(t) + j \mathcal{H}\{x(t)\} \]  

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

\[ x_a(t) = en v(t) e^{j\theta(t)}. \]  

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

\[ x_{SSB}(t) = \text{Re} \left[ en v(t) e^{j(\theta(t)+2\pi \Delta f t)} \right] \]

\[ = x(t) \cos(2\pi \Delta f t) - \mathcal{H}\{x(t)\} \sin(2\pi \Delta f t). \]

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

The frequency shift \( \Delta f \) is a free parameter of the method, which helps to yield a perceptually optimal frequency range of the sonification, e.g., somewhere within the range of 100 Hz and 2 kHz. If \( \Delta f = 0 \), there is no difference to a pure audification. A schematic illustration of the frequency shift is shown in Fig. 2.

Let us consider two scenarios: (1) Assuming a signal bandwidth \( B \) of 10 kHz, a small frequency shift of 100 Hz hardly changes the overall signal, but might make low frequency components of the signal better audible, as the spectrum is now shifted to the range between 100 Hz and 10.1 kHz. Note that in the case of large frequency shifts, the issue of aliasing eventually has to be taken into account. (2) In the second scenario, combining a strong frequency shift with small signal bandwidth results in a very narrow-banded signal which might be problematic from a perceptual point of view. The frequency shift squeezes the original spectrum to a pitch range \( (\Delta f + B)/\Delta f \). For example, if the bandwidth of the primary audification signal is 100 Hz, and the spectrum is shifted by \( \Delta f = 500 \) Hz, the resulting bandwidth is 500–600 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 might be difficult to perceive.

### 3.2 Exponential frequency modulation

Therefore our approach is extended by modulating the frequency shift of the phasor of the analytic signal \( x_a(t) \).

The instantaneous frequency shift of the modulator, \( f_i(t) \), encodes the numeric data values of \( x(t) \) as pitch, i.e. as an exponential function of \( x(t) \), following Flowers’ recommendations:

\[ f_i(t) = 2^{x(t)} \Delta f. \]

The freely choosable parameter \( c \) controls the magnitude of the modulation:

- Setting \( c = 0 \) results in a constant instantaneous frequency of the frequency modulation (FM) which is then independent of the data values \( x(t) \). This results in a pure frequency shift as described in Sect. 3.1.
- Setting \( c = 1 \) leads to a transposition of one octave higher and lower for signal values \( x(t) = \pm 1 \). The value of \( c \) has to be carefully chosen depending on the signal amplitude and bandwidth to prevent aliasing resulting from strong FM sidebands.

Integrating the instantaneous frequency results in the instantaneous phase \( \phi_i(t) \), which serves as a phase modulating term for the analytical signal,

\[ \phi_i(t) = \int_0^t 2\pi \Delta f 2^{x(t)} d\tau. \]

The complete model of Focused Audification is thus defined by:

\[ x_{FA}(t) = \text{Re} \left[ en v(t) e^{j(\theta(t)+\phi_i(t))} \right] \]

\[ = x(t) \cos(\phi_i(t)) - \mathcal{H}\{x(t)\} \sin(\phi_i(t)). \]

The model of FA is controlled by two freely choosable model parameters, \( \Delta f \) and \( c \), that can be set according to the explorative goals of the sonification. Figure 2 illustrates the effect of the parameters with a schematic data set as compared to the absolute hearing threshold.

One issue of FA when dealing with signals of harmonic complexes needs to be discussed. Many physical processes are—at least approximately—periodic. The related signals therefore consist of harmonic partials, and their audification makes use of human audition which groups these frequency components into a single auditory gestalt. In pure audification, frequency ratios and thus the periodicity of harmonic complexes are preserved, resulting in one “sound” with a certain timbre and pitch. In FA on the contrary, the frequency shift destroys the harmonic relationship between the partials and thus the periodicity of the signal. This results in a complex superposition of individual sinusoidal tracks instead of one gestalt with a certain timbre.

### 4 The example of FA of seismological data

As an example with real scientific data we take a file of seismological data from the Incorporated Research Institu-
Fig. 2 A schematic spectrum of a pure audification signal $X(f)$ of bandwidth $B$ is compared to the absolute hearing threshold in a. The frequency shift by $\Delta f$, depicted in b, transposes the spectrum to a more sensitive region of hearing, while narrowing the resulting pitch range. This can be compensated by the frequency modulation controlled by the parameter $c$. The resulting signal $X_{FA}(f)$ is located in a more sensitive region of human hearing and has comparable bandwidth to the original signal $X(f)$.

Fig. 3 Spectrograms of a seismological data set of ca. 5s length with a bandwidth of up to 5kHz (dynamic resolution is limited in the plot), stemming from [13]: a pure audification, b FA, and c FA of section 12–25 seconds (with varying parameter settings).

The sound behaves as an auditory graph, and the former dull glissando event can be explored in detail.

5 Listening test with electrocardiogram data

In this section we present the evaluation of FA in an interactive setting. The main focus of the presented experiments was the adjustment of the free parameters of the model. As data set we chose electrocardiogram (ECG) data: on the one hand, there are well-established, scientifically labeled data sets of ECG data available. On the other hand, these are communicable even to medical laypeople—our test subjects—who are able to categorize these data (this is arguable as, e.g., Ballora et al. [1] have shown that part of their test subjects could achieve 90% correct identification rates with ECG data in a sonification of four different cardiac states). Previous sonification research on ECG signals was conducted by Worrall et al. [29] and Terasawa et al. [24], with a more diagnostic focus.

We performed a pilot test and consecutive experiment, with both quantitative and qualitative analysis, to answer our research questions:
Which are the optimal (preferred and efficient) interindividual parameter settings for the FA of ECG signals?

Which general lessons can be learned using our approach in an interactive, explorative setting?

### 5.1 Experiment design

#### 5.1.1 Choice of data files

We used data from the online MIT-BIH Arrhythmia Database. It is one of the most used ECG databases due to its long and consistent data series \([10,14,17]\), digitized with an average sampling rate of 360 Hz.

One cycle of the basic ECG trace is shown in Fig. 4. Different types of arrhythmias (i.e., irregular heartbeats) have lengthened or shortened intervals within one cycle, or may exhibit an abnormal polarity of the signal part. Furthermore, some arrhythmias appear alternating, where each second or third heartbeat is different.\(^4\)

For our experiment, we chose three types of cardiac states, as labeled in the database: “Paced rhythm”, “Premature ventricular contraction” and “Ventricular trigeminy”. This selection was made following the consultation of an internist with specialization on cardiac insufficiency, ensuring that

\(^4\) In analyzing the ECG signal, the cardiologist uses a template heart beat (averaged over, e.g., a hundred beats) for each of the twelve ECG leads (i.e., taken from 12 positions on the patient’s body), and measures their behavior. This process is often automated today using specialized algorithms to discern different arrhythmias. Still, the expert knowledge of the cardiologists comes into play for border cases, when s/he uses the data plots to explore the signal, as in previous times.

#### 5.1.2 Scenario and task

Our scenario was the usage of FA to monitor ECG data in real-time in an explorative setting. The subjects’ task was to find one optimal parameter set (i.e., for frequency shift and pitch modulation) for all nine sound files, with the goal to

(a) be able to hear differences between the groups A, B, and C (and verbalize their findings) and

(b) have the most amenable sound possible.

NB We did not test if FA is better (e.g., more efficient or its sound more amenable) than other sonification approaches. Developing new methodologies for sonification requires to work also with well-known data sets and environmental conditions of real, scientific, data. For sorting the chosen cardiac states by auditory means a variety of methods could be chosen. The states differ from each other and might even be differentiated by pure audification. In our experiment, subjects were free to choose their preferences from pure audification to FA within a far range of parameters. The efficiency of FA versus, e.g., audification, can be deduced by the theoretical considerations regarding the human hearing in Sect. 2, and would need further testing.

#### 5.1.3 Test conditions

The procedure was repeated in four conditions with different playback rates, see Table 1. We chose these values for the following reasons: we wanted to explore the two free parameters \(\Delta f\) and \(c\), while the playback rate was freely choosable as well. This 3D-space of inter-depending factors is arguably too complex to fully explore within the limited

| Condition | Real-time | Slow | Fast | Adjusted |
|-----------|-----------|------|------|---------|
| \(f_p\) | \(8\) kHz | \(\frac{1}{4}\) kHz | \(40\) kHz | \(3.6\) kHz |

### Table 1

Test conditions of pilot test [only conditions (1)–(3)] and experiment [conditions (1)–(4)] with ECG data, based on a sampling frequency, \(f_s = 360\) Hz

---

[Fig. 4 Normal ECG trace (“sinus rhythm”) comprising the P wave (representing atrial depolarization) and the QT interval (depolarization and repolarization of the ventricles). Source: “SinusRhythmLabels” by Agateller (Anthony Atkielski), licensed under Public Domain via Wikimedia Commons]
time of an experiment. Therefore we fixed three playback speeds in independent conditions. In real-time, the heartbeat of a healthy person varies between 50 and 100 beats per minute, i.e. roughly 1 Hz. Exploring the details of each individual cycle as shown in Fig. 4 requires a slower playback rate that we set to one fourth of real-time as a result of informal tests by the authors. For exploring the macro-structure over several cycles, the acceleration of a factor of 5 was chosen, leading to a rhythm of roughly 5 beats per second, the upper range of the speaking rate measured as vowels per second (see footnote in Sect. 2.1 and [19]). In the pilot test, we only tested conditions (1) to (3). For the full experiment, we added a fourth condition with adapted task: using individual averages, $\Delta f_{\text{mean}}$ and $c_{\text{mean}}$, as calculated over the conditions (1) to (3) for each subject, the subject should set the optimal playback speed.

### 5.1.4 Test procedure

Figure 5 shows the graphical user interface (GUI) of the experiment. For each condition, the subjects were free to choose any sound file and change the parameters as often and as long as they wanted. The experiment was accompanied by an observer, who led through it and collected qualitative data based on an open questionnaire. Questions covered differences between the sound file groups A, B, and C; between the playback speeds, i.e., conditions; and general remarks on the sound quality and the understanding of the mapping (in particular, the correlation between the graph and the sound).

**Test design of the pilot test** The pilot test consisted of two successive rounds that were repeated after a pause that ranged from one hour to 2 days for each subject. The only difference between round one and round two was the use of a different slider design. Our hypothesis was that participants might rely on their visual memory of the slider positions in the first round. Randomly alternating, the subjects were assigned one or the other design first: either the “squared” design (QUAD), where the slider value $x$ behaved like $x = pos^2$, or the square-rooted one (SQRT), where $x = \sqrt{pos}$.

The range for the frequency shift $\Delta f$ was [0, 2000] Hz with varying resolution of maximally 30 Hz for the beginning/end of the slider depending on the slider design (QUAD or SQRT). The pitch modulation $c$ could be chosen within [0, 10] with a maximal step-size of 0.4.

The pilot test showed statistical differences between the two slider designs: this was obviously not intended and due to poor experiment design. Our main hypothesis is that the SQRT design is perceptually counter-intuitive as compared to the exponential dependency between pitch and frequency. The QUAD design is more similar to the psychometric curve of pitch sensitivity [21]. For this reason we re-designed the interaction paradigm of setting the parameters for the main experiment.

**Test design of the experiment** Instead of sliders in the GUI, we used an Apple Mighty Mouse\(^5\) with one miniature track-ball that served as a simple, “endless” slider interface. The subjects did thus not receive any visual or tactile feedback on the position of the parameter setting as compared to its possible range.

The experiment was conducted in one round.

### 5.1.5 Test subjects and time

For the pilot test, 12 test subjects were recruited out of the colleagues of the authors and the authors themselves, i.e., all experienced listeners, but all laypeople in the field of medicine/cardiology (7 out of 12 with a certified hearing loss of less than 15 dB, being part of a trained expert listening panel [9,22]). In the experiment, 14 subjects participated out of which 7 had already taken part in the pilot test (5 of them are part of the trained expert listening panel).

The experiment took roughly 15 min each.

### 5.2 Results

First, we compared results from the two slider designs in the pilot test with the ones of the experiment. Figure 6 shows the 95% confidence ellipses for the parameter settings chosen by the participants. The influence of the slider design on $c$ is not significant for the mean nor the median ($p > 0.43$). However, in the QUAD slider design, participants generally chose smaller values for $\Delta f$ than in the SQRT design ($p < 0.001$). The sequence of the presented slider designs had no significant influence ($p = 0.55$ for $c$ and $p = 0.41$ for $\Delta f$).

Figure 6 also shows that the mean results of the QUAD design

---

\(^5\) https://en.wikipedia.org/wiki/Apple_Mighty_Mouse.
Table 2 Results for aggregated results of pilot test (QUAD) and experiment

| C     | $\Delta f_m$ | $\Delta f_{std}$ | $\Delta f_{conf}$ | $c_m$  | $c_{std}$ | $c_{conf}$ |
|-------|--------------|------------------|-------------------|--------|-----------|-----------|
| (1)   | 279          | 146–534          | 199–391           | 3.11   | $\pm$ 2.74 | $\pm$ 1.43 |
| (2)   | 274          | 151–498          | 200–377           | 1.91   | $\pm$ 1.26 | $\pm$ 0.67 |
| (3)   | 298          | 145–614          | 204–434           | 3.26   | $\pm$ 2.93 | $\pm$ 1.53 |
| all   | 290          | 150–561          | 240–350           | 2.73   | $\pm$ 2.39 | $\pm$ 0.68 |

For settings of parameters $\Delta f$ and $c$ in three conditions and the overall mean of all data are given: mean values ($m$), standard deviations ($std$), and 95% confidence intervals ($conf$). We therefore computed the spectra from the audio files as resulting from the parameter settings for each subject, and compared these spectra to an averaged speech spectrum computed from English, French, and German female and male speakers from EBU SQAM recordings [7]. Figure 8 shows the spectra of sound files of FA for different conditions. Figure 8a depicts all individual spectra for real-time and Fig. 8b “typical” spectra for each condition. A typical spectrum is calculated by taking the logarithmically averaged maximum and average –6dB bandwidth and fitting these values to a Gaussian distribution. It turned out that the individual
Finally, we may draw general conclusions from the qualitative results on the effectiveness of the new method. Preliminary qualitative research of the answers of the test subjects lead to the following conclusions:

- **FA is efficient for categorizing data** Most subjects could verbalize differences between the data categories A, B, and C. As far as the understanding of the authors is concerned (all of us being medical laypeople), these differences correspond well to the specificities of the cardiac arrhythmias as described in the database.

- **FA is flexible in interactive data exploration** In the experiment, we could categorize the participants into two groups: the larger one focused on rhythmic aspects and preferred the fast condition. A smaller group of subjects liked the slow condition better because they were more interested in the details of the modulation within one heartbeat cycle. This finding shows one of the benefits of FA: the interactive, seamless setting of parameters allows to focus on different aspects or scales of a data set. Cardiologists focus both on the behavior within one heartbeat and the general rhythm. Both behaviors have been found and explored by our laypeople listeners.

- **FA provides an acceptable sound** Participants were rather neutral towards the sound quality, many stating that within the context it would be ok (the context had not been defined but assumed by the listeners to be a clinical one). A few participants stated that they would not like to listen to the sound for a longer period. The relationship between the data plot and the sound were reported as clear, even if no participant drew 100% correct conclusions about the underlying mapping. Many participants hypothesized about the data sets, thus they clearly used our approach to explore the data.

Obvious interpretations on the data (e.g., “like a heartbeat”, “again arrhythmies”) were equally mentioned as music metaphors (“a syncopated rhythm”, “strange beat”) or technical ones (“metallic piston noise”, “background noise as in our server room”, “as a remote disco sound”, “chaotic”), and general statements (“unagitated/dull”, “annoying”, “cool/interesting”). It would need a cardiologist to check if the individual findings of the participants could be useful in diagnosis.

### 6 Optimized parameter selection for FA

Concluding from the experiment of Sect. 5 we propose a procedure for selecting the FA parameters \( \Delta f \) and \( c \) for arbitrary data sets. These are optimal in the sense that they adjust the resulting spectrum of the FA as much as possible to the spectrum of average human speech.
1. Normalize the analytical signal to a maximum amplitude of 1.
2. Find relevant events in the signal (either by some a-priori analysis or statistical methods, e.g., an auto-correlation method), and calculate their mean rate $f_{\text{event}}$ related to the sampling rate of the signal (e.g., one event every 100 samples).
3. Choose the playback rate $f_p$ for the FA signal in a way that on the average 3–5 of the relevant events take place within one second, $r_{\text{speech}} \in [3..5]$, to approximate a typical speaking rate,

$$f_p = \frac{r_{\text{speech}}}{f_{\text{event}}} \quad (9)$$

4. Calculate the effective bandwidth $B_{\text{sig}}$ of the analytical signal $x_\text{a}(t)$, i.e. the central second order moment (the variance) of the power spectrum.\(^6\) (Note: if the relevant events constitute only a small part within a long data series, take only the bandwidth of this sequence.)

5. The effective bandwidth of the typical speech spectrum $B_{\text{speech}}$ ranges between $f_{\text{low, speech}} = 125 \text{ Hz}$ and $f_{\text{up, speech}} = 500 \text{ Hz}$, i.e., over 2 octaves, and exhibits a geometric mean frequency $f_{\text{mid, speech}}$ at 250 Hz (see Sect. 5.2). In general,

$$f_{\text{mid}} = \sqrt{f_{\text{up}} \cdot f_{\text{low}}} \quad (10)$$

Based on $B_{\text{speech}}$ we set $\Delta f$ equal to $f_{\text{mid, speech}}$

$$\Delta f = f_{\text{mid, speech}}. \quad (11)$$

By this we achieve a frequency shift of the original signal spectrum to the typical center frequency of the speech spectrum.

Equation 11 is an approximation for signals with negligible bandwidth $B_{\text{sig}}$ as compared to $B_{\text{speech}}$ (as is the case, e.g., for the ECG signals in the experiment above). For the case where $B_{\text{sig}}$ is broadband, ranging from $f_{\text{low, sig}} = f_{\text{sig}} - B_{\text{sig}}/2$ to $f_{\text{up, sig}} = f_{\text{sig}} + B_{\text{sig}}/2$, we need to account for the logarithmic frequency scale and the frequency shift $\Delta f$ has to satisfy the following equation:

$$(\Delta f + f_{\text{low, sig}}) \cdot (\Delta f + f_{\text{up, sig}}) = f_{\text{mid, sig}}^2 \quad (12)$$

Solving the quadratic equation gives as the one physical solution

$$\Delta f = -f_{\text{sig}} \pm \sqrt{f_{\text{mid, sig}}^2 + \frac{B_{\text{sig}}^2}{4}} \quad (13)$$

For the special case of low-frequency signals,

$$\frac{B_{\text{sig}}}{2} \ll f_{\text{mid, speech}}, \quad (14)$$

we may approximate:

$$\Delta f \sim f_{\text{mid, speech}} - f_{\text{sig}}, \quad (15)$$

6. The optimal value of the modulation parameter $c$ is based on the effective speech bandwidth and is set to $c = 1$ for amplitude-normalized narrow-band signals. For broadband signals, the following equation has to be solved for $c$:

$$2c = \frac{f_{\text{up, speech}}}{\Delta f + f_{\text{up, sig}}} = 2 \cdot \frac{f_{\text{mid, speech}}}{\Delta f + f_{\text{up, sig}}} \quad (16)$$

Again, for the special case of low frequency signals (Eq. 14), the optimal value of the modulation parameter $c$ can be approximated\(^7\) by:

$$c \sim 1 - 0.7 \cdot \frac{B_{\text{sig}}}{f_{\text{mid, speech}}} \quad (17)$$

In the case of data sets exhibiting different time regimes of information, i.e. a larger rhythmic structure of events and individual events of interest, we recommend selecting optimal values for $f_p$, $\Delta f$, and $c$ for every event rate found in Step 2 and let the user explore all of them.

With this semi-automatic selection of parameters for FA, the resulting spectrum of the sonification is similar to the one of speech, and thus comfortable for human hearing.

7 Conclusions and outlook

We presented Focused Audification as a method that allows to adjust a sonification between a pure audification and a pitch-based auditory graph. As opposed to pure audification, where only the playback rate can be changed, two more model parameters can be chosen independently. One parameter, $\Delta f$, controls the magnitude of a frequency shift. The second, $c$, sets the excursion of a pitch modulation. The implementation of FA is simple and preserves preferable properties of audification whilst permitting a true “zooming” at any time scale for the interactive exploration of a data set.

\(^6\) Using the identity $x = \int x f(x) dx$ yields $e^{\int\ln x f(x) dx} \cdot \int e^{\ln x f(x) dx} dx$ with the spectral centroid $f_{\text{sig}} = \frac{\int x f(x) dx}{\int f(x) dx}$. For the special case of low-frequency signals, $\frac{B_{\text{sig}}}{2} \ll f_{\text{mid, speech}}$, we may approximate:

$$\Delta f \sim f_{\text{mid, speech}} - f_{\text{sig}}. \quad (15)$$

\(^7\) Using the identity $x = \int x f(x) dx$ yields $e^{\int\ln x f(x) dx} \cdot \int e^{\ln x f(x) dx} dx$ with the spectral centroid $f_{\text{sig}} = \frac{\int x f(x) dx}{\int f(x) dx}$. With taking into account that $\ln(1 + x) \sim x$ for $x \ll 1$, the logarithm can be approximated. This leads to $c = 1 - 1.4427 \cdot \frac{B_{\text{sig}}}{2 \cdot f_{\text{mid, speech}}}$ or $c = 1 - 0.7213 \cdot \frac{B_{\text{sig}}}{f_{\text{mid, speech}}}$, which leads to Eq. 17.
The method has been discussed by the example of a seismo-
logical data set. Preferred and efficient settings for the
model’s free parameters have been explored in an experiment
with ECG data. They appeared to be adjusted in a relatively
narrow region, whose spectrum has maximum energy within
the one of speech. Therefore we concluded on a procedure
to find parameters for FA for any data set in such a way that
their resulting spectrum is similar as much as possible to the
one of speech. Further research has to test the procedure with
different types of data and check the efficiency of FA against
other sonification methods.

Acknowledgements We would like to thank Hanns Holger Rutz for
feedback with the SC implementation, Marian Weger and Franz Zotter
for discussions, and the participants of the experiments for their partic-
ipation in the listening experiments. Thanks to Prof. Dr. F. Skrabal, an
internist with focus on cardiac insufficiency, whom we consulted while
preparing the ECG experiment.

Author contributions RH proposed the idea of Focused Audification,
进一步发展在讨论中与KGV, 谁指导并导
导的实验。MF对re-design for the 2nd
experiment and did the statistical analysis. The optimized parameter
selection was developed by all three authors in discussion. All authors
read and approved the final manuscript.

Availability of data and material Sound examples can be found at http://
phaidra.kug.ac.at/o:92490.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing
interests

Open Access This article is licensed under a Creative Commons
Attribution 4.0 International License, which permits use, sharing, adap-
tation, distribution and reproduction in any medium or format, as
long as you give appropriate credit to the original author(s) and the
source, provide a link to the Creative Commons licence, and indi-
cate if changes were made. The images or other third party material
in this article are included in the article’s Creative Commons licence,
unless indicated otherwise in a credit line to the material. If material
is not included in the article’s Creative Commons licence and your
intended use is not permitted by statutory regulation or exceeds the
permitted use, you will need to obtain permission directly from the copy-
right holder. To view a copy of this licence, visit http://creativcomm
ons.org/licenses/by/4.0/.

Appendix

The software implementation of the proposed method is pre-

dented in the following in order to facilitate its use. We
implemented FA using both MATLAB8 and SuperCollider
(SC).9 MATLAB allows for an analytic use of the method,

8 https://www.mathworks.com/products/matlab.html.
9 https://supercollider.github.io/.

Fig. 9 Synth definition for an FA implemented in SuperCollider (Ver-
sion 3.9.3). The implementation of FA in SC starts from a given buffer
b, with an adjustable playback rate and a start position startpos
from which the buffer read-out starts. The model parameters are called
deltaf and c according to the model definition in Eq. 6. The instan-
taneous frequency fMod is defined, its sine and cosine calculated.
The existing unit generator HilbertFIR returns a two-dimensional array:
hiib[0] contains the primary signal sig (and is multiplied by
the cosine), hiib[1] contains its Hilbert transform (which is multiplied by
the sine). The final output is the difference between those two, according
to Eq. 8

thus we prepared the sound examples and plots discussed in Sect. 4 in MATLAB.
SuperCollider, on the other hand, is
more handy for real-time, interactive use of the method, and
was thus used for the experiment described in Sect. 5. We
present the basic SC Code in Fig. 9.

References

1. Ballora M, Pennycook B, Ivanov PC, Glass L, Goldberger AL
(2004) Heart rate sonification: a new approach to medical diag-
nosis. Leonardo 37(1):41–46
2. Bovermann T, Tünnermann R, Hermann T (2010) Auditory Aug-
mentation. Int J Ambient Comput Intell (IJACI) 2(2):27–41. https://
doi.org/10.4018/jiaci.2010040102
3. Brown L, Brewster S, Ramloll S, Burton R, Riedel B (2003) Design
guidelines for audio presentation of graphs and tables. In: Brazil E,
Shinn-Cunningham B (eds) Proceedings of the international confer-
ence on auditory display (ICAD), pp 284–287. http://eprints.
gla.ac.uk/3196/
4. Davison BK, Walker BN (2007) Sonification sandbox reconstruc-
tion: software standard for auditory graphs
5. de Campo A (2007) Toward a data sonification design space map.
In: Proceedings of the international conference on auditory display
(ICAD)
6. Dombois F, Eckel G (2011) The sonification handbook, Chapter 12—audification. Logos Publishing House, Berlin, pp 237–272
7. European Broadcasting Union (EBU): EBU SQAM CD sound
quality assessment material recordings for subjective tests (2008).
https://tech.ebu.ch/publications/sqamcd. Accessed 16 Dec 2019
8. Flowers JH (2005) Thirteen years of reflection on auditory graph-
ing: promises, pitfalls, and potential new directions
9. Frank M, Sontacchi A (2012) Performance review of an expert
listening panel. In: Fortschritte der Akustik, DAGA. Darmstadt
10. Goldberger A, Amaral L, Glass L, Hausdorff JM, Ivanov P, Mark R,
Mietus J, Moody G, Peng CK, Stanley H (2000) Physiobank, physi-
otoolkit, and physionet: components of a new research resource for
complex physiologic signals. Circulation 101(23):e215–e220.
https://doi.org/10.1161/01.CIR.101.23.e215
11. Hermann T, Hunt A, Neuhoff JG (eds) (2011) The sonification handbook. Logos Publishing House, Berlin
12. Höldrich R, Vogt K (2015) Augmented audification. In: Proceedings of the international conference on auditory display (ICAD), Graz
13. Incorporated Research Institutions for Seismology. http://ds.iris.edu/ds/. Accessed 16 Dec 2019
14. MIT-BIH Arrhythmia Database. http://physionet.org/physiobank/database/mitdb/. Accessed 16 Dec 2019
15. Kildal J (2009) Aspects of auditory perception and cognition for usable display resolution in data sonification. In: Gross T, Gulliksen J, Kotzé P, Oestreicher L, Palanque P, Prates RO, Winckler M (eds) Human–computer interaction—INTERACT 2009. Springer, Berlin, pp 467–470
16. Kramer G (ed) (1994) Auditory Display. Sonification, audification and auditory interfaces. In: Proceedings Volume XVIII, Studies in the Sciences of Complexity. Santa Fe Institute
17. Moody G, Mark R (2001) The impact of the MIT-BIH arrhythmia database. IEEE Eng Med and Biol 20(3):40–45
18. Oppenheim A, Schafer R (1975) Digital Signal Processing. Prentice Hall, Upper Saddle River
19. Pellegroño F, Farinas J, Rouas JL (2004) Automatic estimation of speaking rate in multilingual spontaneous speech. In: International conference on speech prosody, SP–2004, pp 517–520
20. Phillips S, Cabrera A (2019) Sonification workstation. In: Proceedings of the international conference on auditory display (ICAD)
21. Shower EG, Biddulph R (1931) Differential pitch sensitivity of the ear. J Acoust Soc Am 3(2A):275–287. https://doi.org/10.1121/1.1915561
22. Sontacchi A, Pomberger H, Höldrich R (2009) Recruiting and evaluation process of an expert listening panel. In: Fortschritte der Akustik, NAG/DAGA, Rotterdam
23. Speeth S (1961) Seismometer sounds. J Acous Soc Am 33:909–916
24. Terasawa H, Morimoto Y, Matsubara M, Sato A, Ohara M, Kawarasaki M (2015) Guiding auditory attention toward the subtle components in electrocardiography sonification. In: Proceedings of the international conference on auditory display (ICAD)
25. Vézien JM, Ménélas B, Nelson J, Picinali L, Bourdot P, Ammi M, Katz BFG, Burkhardt JM, Pastur L, Lusseyran F (2009) Multisensory VR exploration for computer fluid dynamics in the CoRSAIRe project. Virtual Real 13(4):257. https://doi.org/10.1007/s10055-009-0134-1
26. Vogt K, Frank M, Höldrich R (2015) Effect of augmented audification on perception of higher statistical moments in noise. In: 18th international conference on digital audio effects, DAFx, vol 15
27. Wood SAJ (1973) Speech tempo, vol 9. Working papers, Phonetics Laboratory, Department of General Linguistics, Lund University, pp 99–147
28. Worrall D (2009) Sonification and information—concepts, instruments and techniques. Ph.D. Thesis, University of Canberra
29. Worrall D, Degara N (2014) Detecting components of an ECG signal for sonification. In: Proceedings of the international conference on auditory display (ICAD)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.