Tonality perception of stationary and transient signals

Arne Oetjen
Acoustics Group, Cluster of Excellence "Hearing4all", Carl-von-Ossietzky University, Oldenburg, Germany.

Peter Volk
Acoustics Group, Cluster of Excellence "Hearing4all", Carl-von-Ossietzky University, Oldenburg, Germany; Leopold Kostal GmbH & Co. KG, Lüdenscheid, Germany.

Steven van de Par
Acoustics Group, Cluster of Excellence "Hearing4all", Carl-von-Ossietzky University, Oldenburg, Germany.

Summary

For environmental noise, especially for machine noise, annoyance tends to increase with increasing tonality. Tonality also represents an important feature in the overall sound character. A psychoacoustical experiment was set up for determining the degree of perceived tonality for various masker types. For this purpose points of equal tonality were measured with an adaptive procedure adapting the level of a pink noise masking a fixed tonal signal. The results, for pure tones with several SNRs in different masker situations showed that the perceived tonality does not depend on individual detection thresholds for normal-hearing listeners. Instead it was found that, in line with DIN45681 (2005) [1], the tonality strongly depends on the SNR within a critical band around the tonal component. In a further experiment the tonality was measured for tonal components changing rapidly in frequency. Contrary to what is perceived in informal listening, the DIN45681 [1] predicts a very strong decrease in tonality. As a result tonality analysis may become problematic for example for combustion engines equipped with turbochargers. As a first approach to investigating the tonal perception of time-variant sounds, detection thresholds of sinusoidal logarithmic sweeps masked with pink noise were determined in dependence of sweep duration and frequency range. Results show that listeners can detect sinusoidal sweeps, but that thresholds increase to some degree with increasing sweep rate. To conclude a new algorithm for tonality detection based on an auditory filter bank approach is introduced. This algorithm is capable of detecting fast frequency changes as they might occur in many natural sounds.

PACS no. 43.66.Lj, 43.66.Ba, 43.66.Ki

1. Introduction

Quantifying the tonality of an environmental sound plays an important role for example in product development processes involving sound analysis and sound design. Tonality can be expressed as the strength of tonal content in the overall sound. Tonality calculation algorithms like in the German industry standard DIN45681 [1] model this subjective sound impression based on the signal-to-noise ratio (SNR) of a sinusoidal tone and the noise floor measured within a critical band. This principle is based on psychoacoustical experiments which typically involved stationary pure tones in broadband noise. Since individual detection thresholds for these stimuli show quite large differences the question arises how the individual tonality perception of such stimuli is connected to the detection SNR. In order to create a better understanding of the mechanisms underlying tonality perception, tones were presented in four noise types that differ in their spectral composition. In a second experiment detection thresholds for pure tone sweeps with different duration and sweep rate are measured. A new model approach being able to detect stationary as well as highly transient tonal components will be presented.

2. Apparatus and Stimuli

The test stimuli were chosen such that they resemble different masking situations with a sinusoid partially masked by notched noise, and by far and near band-pass filtered noise signals. In a fourth condition, narrow-band low-noise noise (LNN) created the tonal impression instead of a sinusoid. The spectra of these sounds are illustrated in Figure 1.

15 normal-hearing listeners participated in the psychoacoustical experiments, their age ranged from 22 to 30 years. All participants were experienced in listening experiments. All stimuli were generated digitally and presented via headphones (Sennheiser HD 650) in a single-walled sound-attenuating listening booth.

Each sound was presented in a 1 s interval with 25 ms of $\cos^2$ ramps for fading in and out. The tonal parts of the signal were 0.9 s long and presented temporally centered inside the signals. The noise components were played back at a level of 70 dB (A) SPL.

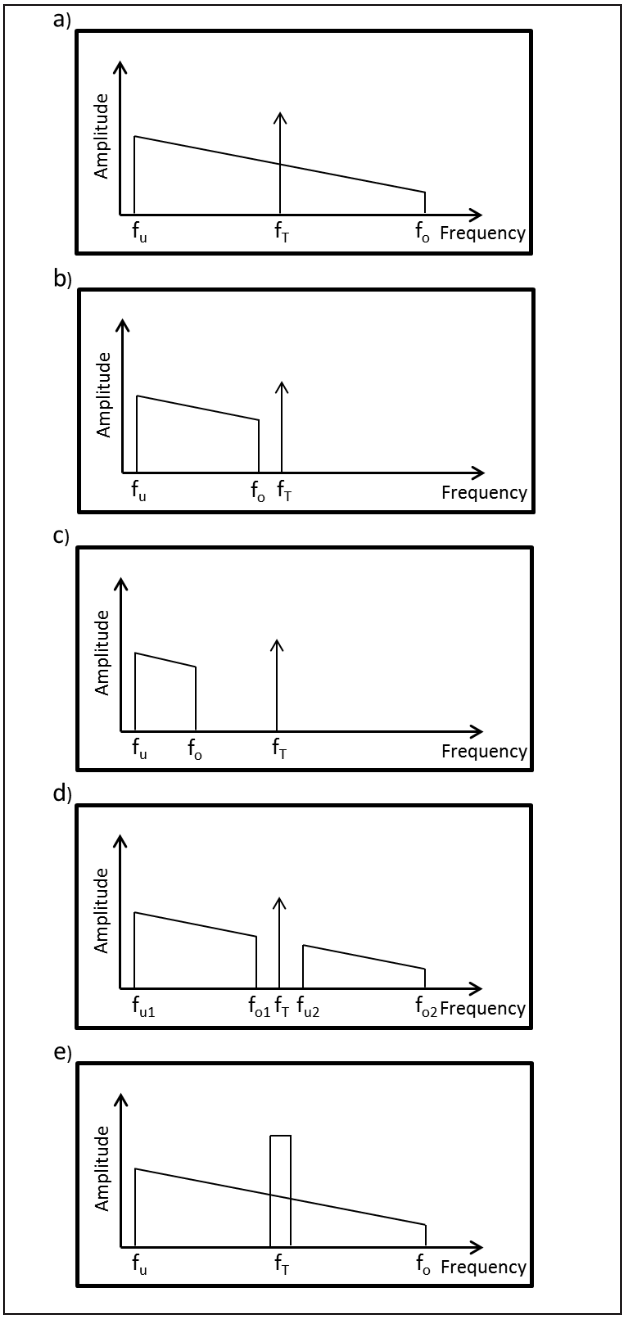

Figure 1. Schematic representation of the stimuli spectra. For all stimuli the tone frequency f_T was 1 kHz and the lower cut-off frequency of the pink noise f_u was 100 Hz. For the reference condition a) the upper cut-off frequency of the noise, f_o, was 10 kHz. For the tone near noise condition b) (TnN) f_o was 0.5 ERB below the tone frequency f_T, for the tone far from noise condition c) (TffN) f_o was 2 ERB below the tone frequency. For the tone in notched noise condition d) (TiNN) the lower cut-off frequency of the notch f_{o1} was 0.5 ERB below the tone frequency and the upper cut-off frequency of the notch f_{u2} was 0.5 ERB above the tone frequency. f_{u1} and f_{o2} was the same as f_u and f_o in the reference condition. For the low-noise noise in condition e) (LNNiN) the noise was the same as in the reference condition, the low-noise noise had a bandwidth of 50 Hz.

3. Experiments and results

3.1. Detection Thresholds

The lowest possible tonality for a sound is determined by the detection level of the tonal component. Individual detection thresholds for all five tone-in-noise conditions (see Figure 1) were determined by an adaptive 3-alternative-forced-choice procedure. The tone-in-noise signal was presented in one randomly chosen interval while the masking noise alone was presented in the two other intervals. The listener's task was to choose the interval containing the tonal component. Mean values and standard deviations as well as maximum and minimum thresholds across all listeners are shown in Figure 2.

The individual detection threshold levels for the tone differed up to 10 dB for the condition where the tone frequency was placed 2 ERBs above the upper cut-off frequency of the noise and around 5 dB for all other masking conditions. These high individual differences in detecting tonal components at low SNRs pose the question whether tonality perception can be generalized for a large population or if the individual differences in detection thresholds are reflected in a tonality perception differing across listeners by in the same manner as the detection thresholds for tonal components.

3.2. Tonality matching for stationary signals

In a next step the tonality for different SNR conditions of the different stimuli conditions was determined. This was carried out in a tonality matching experiment implemented in an adaptive matching procedure. For each stimulus condition the listeners were

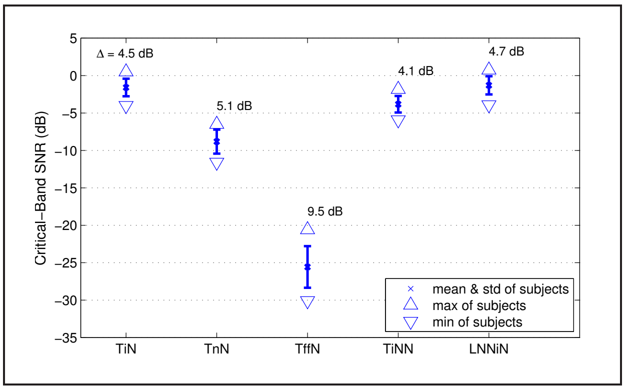

Figure 2. Mean detection thresholds and standard deviations for tones in noise expressed as the SNR in one critical frequency band (y-axis) for the five tone-in-noise conditions described in Figure 1 (x-axis). Maximum and minimum individual thresholds are expressed as triangles pointing up- or downwards, the difference between these thresholds is denoted by the numeric values.

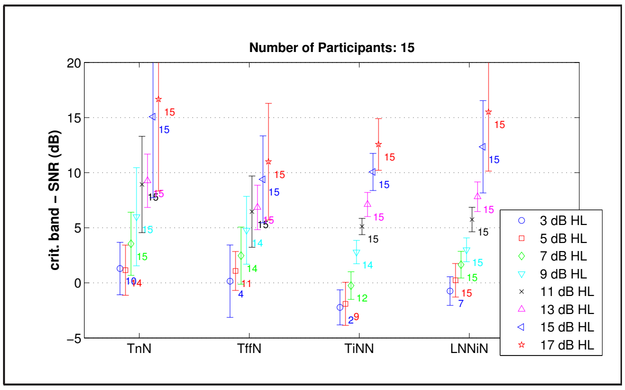

Figure 4. Mean values and standard deviations of SNRs in one critical band of the reference signal at equal tonality as the test stimulus (y-axis) for the four test stimuli b)-e) described in Figure 1 (x-axis). The different SNR conditions ranging from 3 to 17 dB above the lowest individual detection threshold (see Figure 2) of the test stimuli are shown in different colors (see legend). The numeric values on the right side of the single data points denote the number of listeners that were able to detect the tone at the given SNR.

asked to compare the tonality of a reference signal, in this case the band-limited pink noise with a sinusoid as described in Figure 1, with the tonality of a test signal. The test signals consisted of the four other tone-in-noise conditions shown in Figure 1 with different SNR levels. If the listener rated the tonality of the reference signal higher than the tonality of the test signal the noise level of the reference signal was increased before the next presentation of the signal beginning with a step size of 5 dB which was finally reduced to 1 dB. The point of subjective equality for the tonality was then calculated as the mean value of the last six reversal points in the adaptive procedure. The single comparisons for all test conditions were presented interleaved with each other in order to avoid unwanted responses strategies to a single condition by the listeners.
The lowest SNR values used in test signals were 3 dB higher than the lowest detection thresholds obtained in the experiment described in section 3.1, i.e. 3 dB above the downward pointing triangles in Figure 2. The next SNR condition were respectively increased by 2 dB, in total eight conditions, ranging between 3 to 17 dB above the lowest individual detection threshold, for each of the four masker conditions were used in the comparison experiment. As some listeners did not detect tone at a threshold 3 dB above the lowest individual threshold or at higher SNRs they were not asked to perform the experiment for these SNR conditions.
One of the issues of this study was to find out whether tonality perception correlates to the tonal level increment above the individual detection threshold or whether it better correlates to the plain SNR in a critical band as suggested in the tonality model in DIN45681 (2005) [1]. In Figure 3 the individual SNRs of the matched reference signals are shown for the tone in notched noise condition, in the left panel as SNRs in one critical frequency band relative to the individual detection threshold, and in the right panel as absolute SNRs in one critical band. Obviously the individual results differ much less if they are expressed as absolute SNRs, which points to the tonality perception being independent of the individual detection threshold. The same behavior was found for the three other masking conditions. Therefore all mean values are taken from the absolute SNRs for equal tonality of reference and test signal, these data are shown in Figure 4.
The results show that the SNRs of the reference stimulus at equal tonality as the test stimuli approximately have the same dynamic range like the SNRs of the test stimuli themselves (14 dB). One interesting observation may be the compressive behavior of the adjusted SNRs for low-noise noise (LNNiN) at low SNRs but this has not been subject of further research yet.

3.3. Tone detection for transient signals

Although in the previously experiments only the tonality of stationary signals was tested, tonal components in natural sounds such as for example machinery noise often occur with varying frequencies. Especially rapid frequency changes are not detected as tonal components by the algorithm proposed in DIN45681 (2005) [1]. This may be problematic when, for example, trying to detect tones emitted by turbocharged engines during the acceleration process at a low rotational engine speed.
Previous studies showed a dependency of detection thresholds on the signal duration [2]. Based upon these findings the influence of the rate of a frequency change on detection thresholds was investigated by

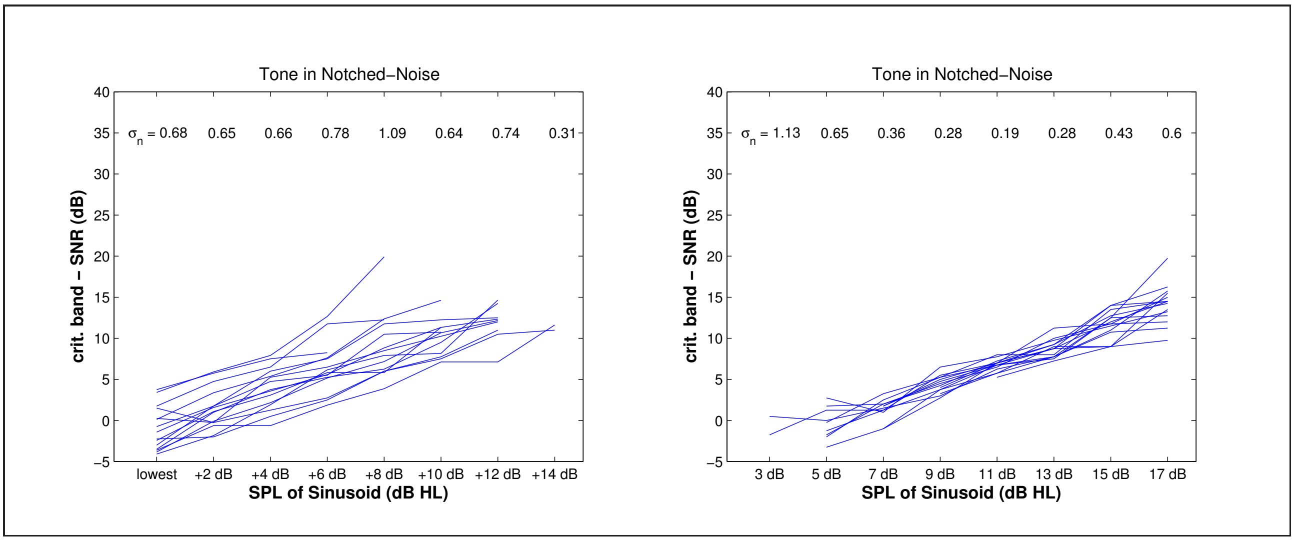

Figure 3. Individual SNRs of the reference signal at equal tonality as the test signal in one critical band (y-axis) relative to the individual detection thresholds (left panel) and absolute SNRs (right panel) for different SNR conditions of the tone in notched noise (TiNN) test signal (x-axis). The different lines represent the single individual points of equal tonality, the corresponding standard deviations across all listeners for each SNR condition are shown as numeric values inside the figures.

usage of the same measuring equipment, listeners collective, and the same adaptive procedure described in section 3.1. As a masker the same pink noise as the one in the reference stimulus shown in Figure 1 was presented for 1 s at a level of 70 dB(A). The duration of the tonal components was chosen as 0.1, 0.25 and 0.5 s, these stimuli were presented temporally centered inside the target interval. The tonal components were faded in and out with raised cosine ramps with a length of $^1/_{80}$ of the tone duration. The signals were logarithmic sweeps centered around 1 kHz and sweeping through a frequency range of one or two octaves. As a reference condition, detection thresholds of a pure tone with the same duration as the sweeps were determined additionally. Each participant completed the experiment three times. Median values and 25 and 75 % percentiles across all listeners are shown in Figure 5.

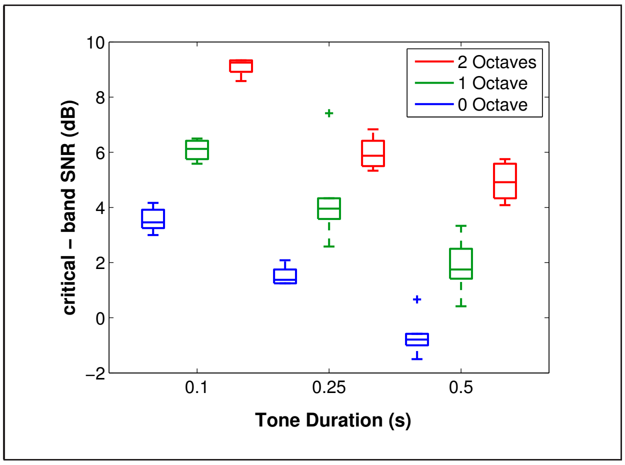

Figure 5. Box plots indicating the median values and the 25 and 75 % percentiles across all listeners of the SNRs in one critical band (y-axis) at detection threshold for logarithmic sinusoidal sweeps centered around 1 kHz with different sweeping duration (x-axis). The crossed frequency range Δ_f was zero octaves (blue), i.e. stationary pure tones, one octave (green) and two octaves (red).

From the results, a dependency of detection thresholds on signal duration is clearly visible; the thresholds for pure tones decrease by about 4.5 dB for signal durations of 0.1 and 0.5 s. The thresholds for the sweeps over one octave are approximately 2.5 dB higher than the ones for pure tones, the thresholds for the two-octave sweeps are shifted by about 6 dB. Since the sweep rate changes with the signal duration while the effect of changing the stimulus duration is the same as for stationary tones, one may conclude that the detection of transient tonal signals may not depend on the sweep rate but on the sweep range giving a constant offset in detection thresholds.

4. A new model approach for tonality detection

As the tonality calculation based on audio spectra proposed in DIN45681 (2005) [1] does not detect transient tonal components, especially those with rapid frequency changes, an alternative model approach was developed. Instead of a more technical approach of the DIN45681 (2005), a model based on the mechanisms of human hearing using a gammatone filterbank [3] was preferred. All stages of the hearing model were

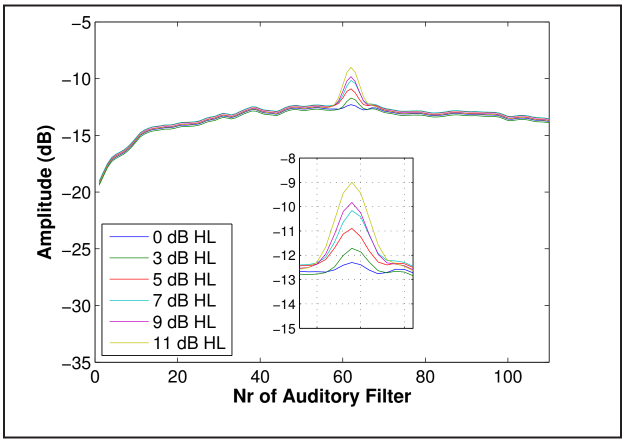

Figure 6. Mean values from the output of the 0 Hz modulation filter (y-axis) for each auditory filter (x-axis). The analyzed stimulus was low-noise noise in noise at different SNRs (different colors, see legend). In the small picture, the sensitivity to different levels of tonality can be seen.

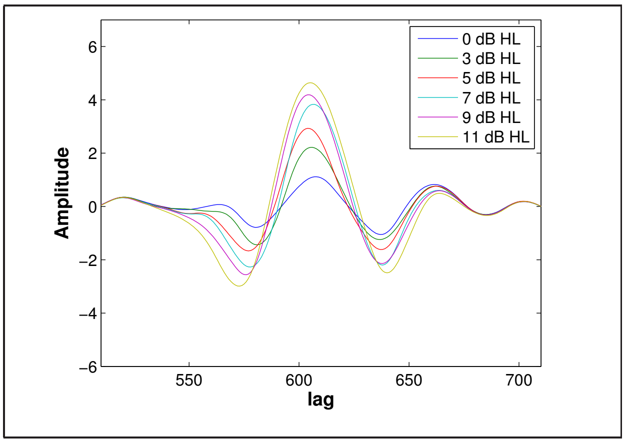

Figure 7. Summed 0 Hz cross co-variance for LNN in Noise at different SNRs (colors)

provided by the "Auditory Modeling Toolbox" [4]. Based on the concept that the modulation spectra of noise get broader as the bandwidth of the noise increases [5] the approach was to extract the DC component of the modulation spectrum as a basis for the tonality detection, similar to the Envelope-Power-Spectrum-Model (EPSM) [6] but working with mean values instead of the energy detector. Among other modifications to the filterbank model, the number of auditory filters was increased and the individual bandwidths of the filters were slightly reduced. This approach resulted in a single value related to the output level of the 0 Hz modulation filter for each channel of the auditory filterbank. An example for such a representation is shown in Figure 6 for low-noise noise in noise at different SNRs.
In order to obtain a tonality detector that would also clearly react on the distinctiveness of tonal components, some excitation patterns obtained with pure sinusoids of different frequencies are stored. At a second step the cross co-variance between the excitation pattern of the current input signal and the stored excitation patterns is calculated. The single co-variance patterns obtained for the different training patterns are summed up in a next step making the approach very robust against noise which gets canceled by this averaging process. An example for low-noise noise in noise at different SNRs in such a representation is shown in Figure 7.
The peak height of the patterns in this summed co-variance representation can now be regarded as the calculated tonaltity. A transformation to the more common dB values has not been implemented yet, but these "arbitrary" values already show a very good behavior as a detector for tonal frequencies and tonality itself. In Figure 8 the performance on the sweep stimuli from the previous section is shown. The algorithm shows a very high capability of detecting transient and stationary tonal components which, in the case of transient stimuli, are not detected as tonal components by the established approach [1].
One main motivation for the development of this new approach was among others the very rapidly sweeping tonal components of turbocharged car engines. Due to the lack of a time-averaging process, the new method should now be capable of detecting these kinds of tonal components very well. In Figure 9 the tonality spectrogram of a sound is shown. It is recorded while a passenger car equipped with a turbocharged Diesel engine drives past a microphone during acceleration process such that the turbocharger starting to rotate in a clearly audible manner. This tonal component is spanning a frequency range of about 10 kHz during 1 s of the turbocharger's acceleration time. During this process a lot of other sources are emitting tonal and noisy sound components but the algorithm is still performing quite satisfactory.

5. Summary and Conclusion

In this study some new data for tonality behavior in different masking situations and differing distinctiveness of tonal components were collected. This also includes tonality ratings for masked low-noise noise which showed that listeners clearly attribute tonality to the sound sensation of narrow-band low-noise noise. The detection of sound components with varying frequency seems to depend on the signal duration and the crossed frequency range, the sweep rate itself does not seem to influence the detection thresholds.
Furthermore, a new model approach was introduced being able to detect tonal components for tonal noise conditions as well as for components with very fast frequency changes. This calculation method also shows results highly related to the human tonality perception for natural sounds such as machinery noise, even if many different sound sources are contributing tonal components and noise at the same time.

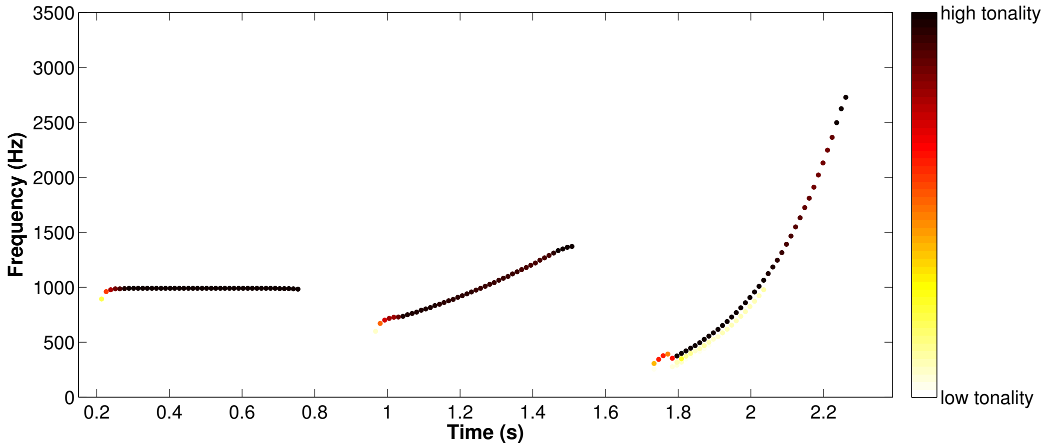

Figure 8. Tonality analysis spectrogram for logarithmic sweeps and a pure tone (left) with a duration of 0.5 s each (see Figure 5). The sweep range was one octave (middle) and two octaves (right). Darker colors indicate higher tonality.

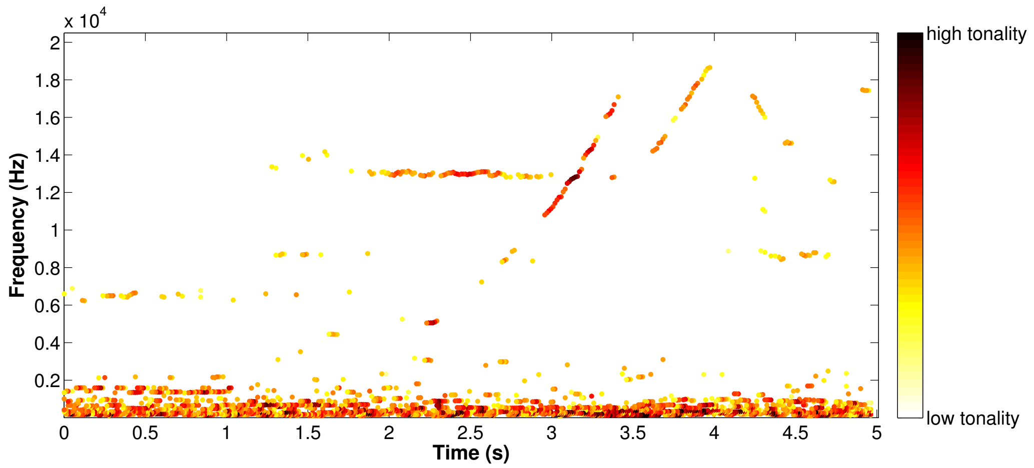

Figure 9. Tonality analysis spectrogram for an accelerating passenger car equipped with a turbocharged diesel engine passing a microphone. Between 2.5 and 3.5 s the transient tonal sound component of the turbocharger starting to rotate is visualized by the calculation algorithm. Darker colors indicate higher tonality.

References

[1] DIN45681: Akustik - Bestimmung der Tonhaltigkeit von Geräuschen und Ermittlung eines Tonzuschlages für die Beurteilung von Geräuschimmissionen. DIN Deutsches Institut für Normung e.V. (2005)

[2] B.C.J. Moore: An Introduction to the Psychology of Hearing. Brill (2013)

[3] V. Hohmann: Frequency analasys and synthesis using a Gammatone filterbank. Acta Acustica/Acustica (2002) 433-442.

[4] P. Søndergaard and P. Majdak: The Auditory Modeling Toolbox. The Technology of Binaural Listening (2013) 33-56.

[5] J. Lawson and G. Uhlenbeck: Threshold Signals. Dover Publications, Inc. (1950).

[6] S. Ewert and T. Dau: Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc. Am. 108(3) (2000) 1181-1196.