ICA 2019 AACHEN

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Progress in Tonality Calculation

Julian BECKER[1]; Roland SOTTEK[2]; Thiago LOBATO[3]

[1,2,3] HEAD acoustics GmbH, Germany

ABSTRACT

The perception of tonal sounds is one of the most important psychoacoustic sensations for product sound quality and environmental noise. In the past, a lot of work has been carried out to quantify tonal sound events automatically. This work resulted in several methods for quantifying tonalities, such as Tone-to-Noise Ratio, Prominence Ratio or the Psychoacoustic Tonality, which was standardized in the 15th Edition of the ECMA-74 standard. The Psychoacoustic Tonality is based on a hearing model which emulates the processing of human hearing. Thus, it is able to predict human perception better than other methods which are only partly based on psychoacoustics.

Since the publication of the Psychoacoustic Tonality, the algorithm is constantly being improved. In this paper, the latest progress in the calculation of the Psychoacoustic Tonality is presented. The new developments include improvements for low frequencies, identification of tonal components and a higher frequency resolution of the detected tonalities.

These improvements will likely be included in future editions of the ECMA-74 standard.

1. INTRODUCTION

Technical and natural sounds frequently contain prominent tonal components. These sounds are often either produced by periodicity, for example by a rotating device, or by narrowband noises which can for example be generated by air flow.

Tonal components are perceived very prominently by a human listener and thus significantly influence the individual perception and evaluation of a sound event. Tonal sounds significantly increase annoyance, if they are perceived as unwanted. Hence, the quantification of tonal sounds has been an important topic for a long time. The topic currently gains even more attention due to the increasing importance of electric vehicles. While these vehicles produce less overall noise, electric motors do produce a rather tonal sound.

Several attempts have been made in the past for the automatic quantification of tonalities. Methods like the Prominence Ratio (PR) [1] or the Tone-to-Noise Ratio (TNR) [2] have already been in use for a while, but they often lead to unsatisfactory results, because they do not consider human perception. As a consequence, both methods produce implausible results in certain scenarios. For example, the same signals with different sound pressure levels lead to the same result when using these methods even though the perception of tonality has a level-dependence.

Research results show a strong correlation between tonality perception and the partial loudness of tonal sound components [3]-[5]. Recently a new method, the Psychoacoustic Tonality, has been published in the 15th edition of the ECMA-74 standard [6]. This method is based on a psychoacoustic hearing model and thus takes into account several psychoacoustic effects that are not considered in PR and TNR. The applicability of the model was investigated for technical sounds and compared to established methods [7,8].

Besides the quantification of tonality, recent research also deals with the quantification of tonal annoyance [9]. The ECMA-74 tonality can also be modified to describe this sensation [10].

While the Psychoacoustic Tonality shows a better performance than other established methods [11],

[1] julian.becker@head-acoustics.de
[2] roland.sottek@head-acoustics.de
[3] thiago.lobato@head-acoustics.de

EAA ICA DEGA

it is constantly being revised to further improve the results and to include additional features. These improvements are presented in this paper. Recently, it was shown that the ECMA-74 Psychoacoustic Tonality tends to overestimate tonalities of tonal components at low frequencies [10]. To avoid this effect, the outer and middle ear filtering of the algorithm was revised. New or improved features are introduced to the algorithm by identifying tonal components and increasing the frequency resolution of the detected tonalities. If not otherwise mentioned, ECMA-74 refers to the 15th Edition of the ECMA-74 standard [6] throughout this document.

2. PSYCHOACOUSTIC HEARING MODEL BASED ON ECMA-74

The basis for the tonality calculation is a model of the human hearing which transforms sound pressure to perceived loudness. This model was first described in [12] and later improved further. In the 15th edition of the ECMA-74 standard [6] it is described in Annex F and used as basis for further calculation of psychoacoustic analyses. In this section the hearing model according to ECMA-74 is first described and improvements which were recently applied to the model are presented.

Figure 1 shows the structure of the hearing model. The sound pressure input signal $p(n)$ is processed according to the signal processing of the human hearing. Finally, the model results in the specific loudness $N'(z)$. The processing steps of the hearing model are briefly discussed in the following and recent improvements to the algorithm are described. Specific information about the exact implementation can be found in the ECMA-74 standard [6]. Since the definition of the critical bands in the hearing model differs slightly from the commonly used definition by Zwicker [13], the unit $\mathrm{bark}_{\mathrm{HMS}}$ (bark according to the hearing model of Sottek) is used in the following. Correspondingly the unit $\mathrm{sone}_{\mathrm{HMS}}$ is used for the loudness calculated by the hearing model.

The processing starts with outer and middle ear filtering. In ECMA-74, this filter is defined such that the resulting loudness of the hearing model fits well to the equal loudness contours (ELCs) of ISO 226:1987 for frequencies below 1 kHz and to the ELCs of ISO 226:2003 for higher frequencies.

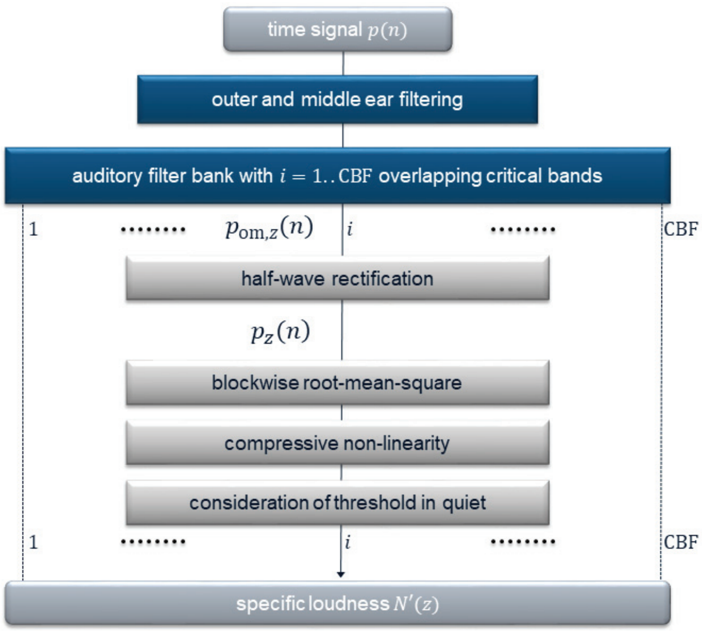

Figure 1 – Basic hearing model structure

Since the results of ECMA-74 tend to overestimate tonalities at low frequencies, this filter was revised such that the results also match the ELCs of ISO 226:2003 for frequencies below 1 kHz. Additionally, a peak at frequencies above 10 kHz was removed since it artificially added tonal effects to some signals. Also, the filter was designed such that it can be described as two serial filters corresponding to the influence of the outer ear and of the middle/inner ear. The outer ear filter can be divided into a filter considering the diffuse field influence and a filter considering additional effects in free field. With this definition, it is possible to differentiate between free field and diffuse field conditions. In free field conditions, the complete filters need to be considered, while in diffuse field conditions only the diffuse field part of the outer ear needs to be used together with the filter of the inner/middle ear.

Figure 2 shows the original filter as described in ECMA-74 (left) and the modified filter with the two serial filters corresponding to the outer ear and to the middle/inner ear for free field condition.

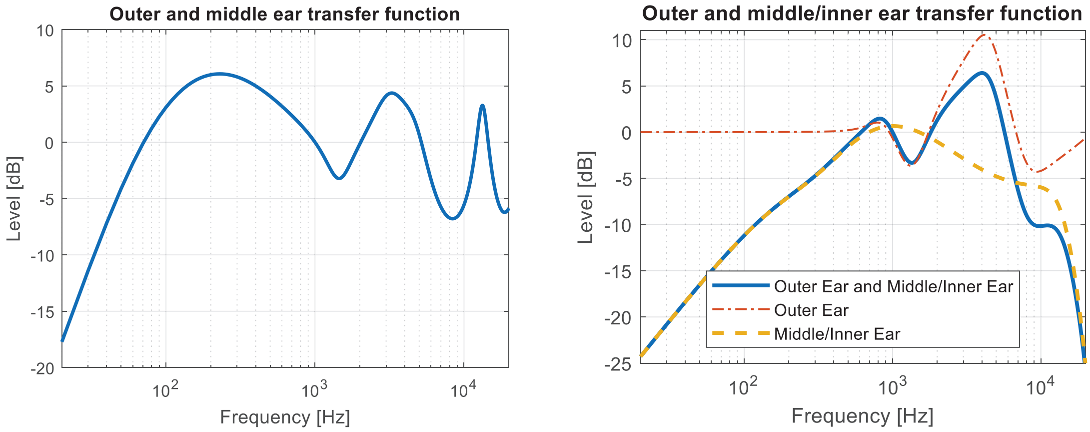

Figure 2 – Left: Outer and middle ear transfer function according to ECMA-74,
Right: Revised outer and middle/inner ear transfer function

In the next step, the signal is processed by an auditory filter bank consisting of 53 overlapping filters corresponding to critical bands from $z = 0.5$ to $z = 26.5$ with a step size of $\Delta z = 0.5$. This step together with the subsequent half-wave rectification remains unchanged and is calculated exactly as described in ECMA-74. The resulting 53 rectified band-pass signals $p_z(n)$ are later used as basis for the calculation of the tonality.

To calculate the specific loudness, blockwise root-mean square (RMS) values are taken from $p_z(n)$ and transformed nonlinearly to result in specific loudness without consideration of the threshold in quiet. These steps are calculated unmodified according to ECMA-74. For the calculation of the blockwise root mean square values, a frequency-dependent block length $s_\mathrm{b}(z)$ is used. The block length is 8192 for the lowest critical bands and is decreased stepwise by a factor of 2 until it reaches a length of 1024 samples.

Finally, the effect of the threshold of hearing is applied to the loudness values. The function describing the specific loudness threshold needs to be modified to fit to the outer and middle ear filters.

Figure 3 shows the specific loudness thresholds according to ECMA-74 (left) and the modified one (right).

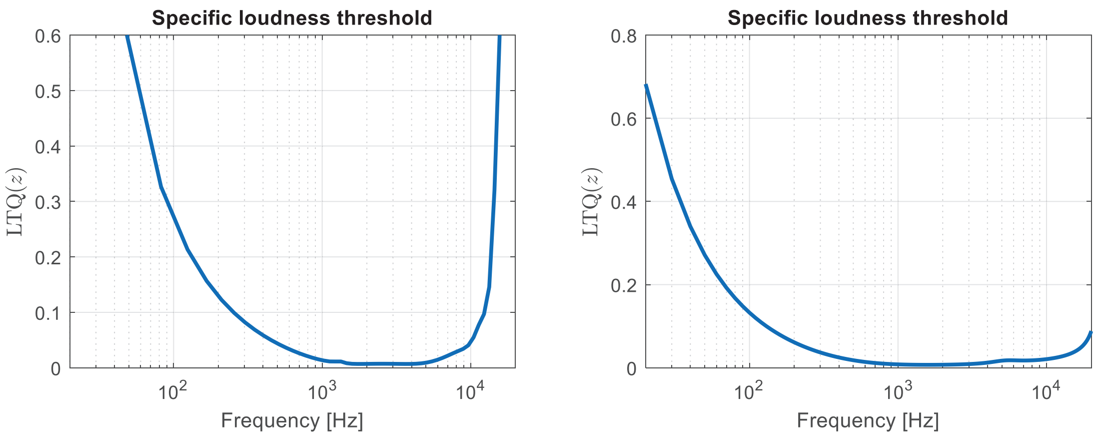

Figure 3 – Left: Loudness threshold according to ECMA-74, Right: Revised loudness threshold

The results of this calculation are 53 specific loudness values $N'(z)$ for each time block, which are used as basis for the tonality calculation process.

3. CALCULATION OF TONALITY

The calculation of tonality is based on the rectified band-pass signals $p_z(n)$ which were calculated in the hearing model. Additionally, the final result of the hearing model, the specific loudness $N'(z)$ is used. The basic structure of the tonality calculation is shown in Figure 4.

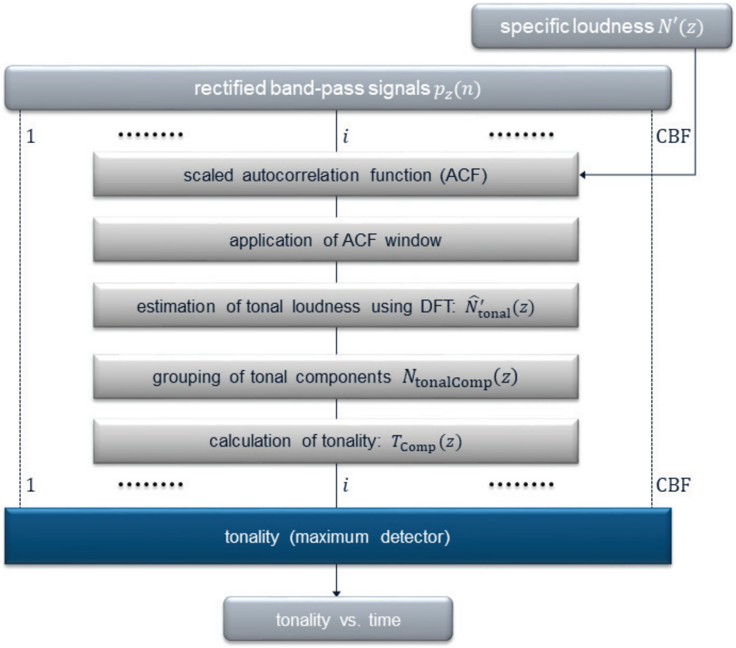

Figure 4 – Basic structure of the tonality calculation

3.1 Estimation of Tonal Loudness.

Tonality is calculated from the tonal loudness of the signal. The estimation of tonal loudness is done similarly to ECMA-74 but with minor modifications. The first step to estimate the tonal loudness is to calculate blockwise scaled autocorrelation functions (ACF). The block size $s_{\mathrm{b}}(z)$ which was used for the calculation of the RMS values in the hearing model is used. The calculation of the autocorrelation function is done according to ECMA-74.

In ECMA-74, ACFs of neighboring bands and of neighboring blocks are averaged to reduce noise effects. This averaging step is skipped since noise reduction is inherently performed in a new grouping step which will be described later.

Next, a window is applied to the ACFs as described in ECMA-74. Since signal energy of noisy signal parts only appears for low lags of the ACF, the window is placed such that lower lags are ignored. Tonal loudness is then estimated by first removing the mean of the windowed ACF. The windowed ACF with removed mean is analyzed by a Discrete Fourier Transform (DFT). In ECMA-74, a $2s_{\mathrm{b}}(z)$-point DFT is used, resulting in different frequency resolutions depending on the band z, with the highest frequency resolution for the lowest bands. To obtain a constantly high frequency resolution, a 16384-point DFT is now used independently of the block size. This size corresponds to the $2s_{\mathrm{b}}(z)$-point DFT for the lowest frequency bands but results in a higher resolution for higher bands. Tonal loudness is then estimated as the maximum of this spectrum. In ECMA-74, a frequency range is defined to limit the maximum search. This frequency range is no longer used in order to provide consistent results for the entire audible frequency range between (specific) loudness and (specific) tonal loudness. Apart from this, the tonal loudness is estimated as described in ECMA-74. Additionally, the frequency at which the maximum occurs is saved as the frequency corresponding to the tonal sound.

Since the block size used for the ACF differs for different frequency bands, the temporal resolution of the blocks is also different. Thus, the estimated tonal loudness and the corresponding frequencies are resampled to a common time basis as explained in ECMA-74. With this resampling step, all results have a common time index l. Since the next steps are independent of this time index, it will be neglected for the description of the algorithm and only be reintroduced for the description of the time-dependent tonality. This first estimate of the tonal loudness is named $\hat{N}'_{\mathrm{tonal}}(z)$ and the corresponding frequencies $f_t(z)$. The resampling is also applied to the specific loudness of the signal $N'(z)$, resulting in the estimated loudness of the signal $\hat{N}'_{\text{signal}}(z)$.

3.2 Low-pass Filtering

In ECMA-74, the first estimate of the tonal loudness is refined by low-pass filtering and a noise reduction step by applying a sigmoid function depending on a SNR estimate. The low-pass filtering is performed in the same way as described in ECMA-74, resulting in the final estimate of the specific tonal loudness $N'_{\text{tonal}}(z)$. The low-pass filtering is also applied to $\hat{N}'_{\text{signal}}(z)$, resulting in the final estimate of the loudness of the signal $N'_{\text{signal}}(z)$.

The application of the sigmoid function is skipped because a new grouping step is introduced which will be described in the next section. This grouping step inherently performs noise reduction.

3.3 Grouping of Tonal Components

In the grouping step, the specific tonal loudness $N'_{\text{tonal}}(z)$ and the corresponding high-resolution frequency $f_t(z)$ are used to group the specific loudness of tonal components into the total loudness of those components. To describe the grouping step, $N'_{\text{tonal}}(z)$ and $f_t(z)$ are described as a function of an index i ranging from 1 to 53 rather than a function of the critical band number z, i.e. $N'_{\text{tonal}}(i)$ and $f_t(i)$ are used. The resulting tonal loudness values of the components are stored in $N_{\text{tonalComp}}(i)$, which is first initialized with zeros for each index i and then filled with the corresponding tonal loudness values in an iterative process as described below. $N_{\text{noiseComp}}(i)$ is the corresponding variable for the noise loudness in the bands of the tonal components.

First, the frequencies corresponding to the tonal content of each band, $f_t(i)$, are transformed to bark_{HMS} values $z_t(i)$ according to Equation (F.6) of the ECMA-74 standard (15th Edition).

Next, the iterative process starts. For the first iteration, a set of indices I is defined, which first contains all indices i which belong to a defined frequency range:

$$I = \{i \mid z_t(i) \geq z_{\min} \wedge z_t(i) < z_{\max}\}, \tag{1}$$

where $z_{\min}$ and $z_{\max}$ are critical band values corresponding to the range of human hearing. Values corresponding to a lower limit of 16 Hz and an upper limit of 20 kHz are used. The grouping is only performed if the overall tonal loudness of all $i \in I$ exceeds 0.01 sone_{HMS}:

$$\sum_{i \in I} N'_{\text{tonal}}(i) \cdot \triangle z > 0.01 \ \text{sone}_{\text{HMS}}. \tag{2}$$

Next, new sets of indices are defined for each index $j \in I$, containing all indices with a critical band value $z_t(i)$ which is close to the value $z_t(j)$ of the specific band index j:

$$C_j = \{i \in I \mid z_t(i) \geq z_t(j) \wedge z_t(i) < z_t(j) + \triangle \text{b}\}, \tag{3}$$

where $\triangle \text{b}$ is the maximum allowed deviation. A value of $\triangle \text{b} = 0.2$ bark_{HMS} is used.

Each set of indices C_j stands for one potential group corresponding to one tonal component. Only the group corresponding to the highest tonal loudness is considered as a tonal component in this iteration step. This is the set $C_{j_{\max}}$, corresponding to the index $j_{\max}$ maximizing the tonal loudness of the group:

$$j_{\max} = \arg\max_{j \in I} \sum_{i \in C_j} N'_{\text{tonal}}(i) \cdot \triangle z \tag{4}$$

The set $C_{j_{\max}}$ is only accepted as a tonal component if it contains at least 3 elements for $j_{\max} > 2$, resp. 2 elements for $j_{\max} \leq 2$). The reason for this limitation is that a tonal component will usually appear at least in the directly neighboring bands. If it does not appear in the neighboring bands, it can be assumed that the high loudness originates from a noisy sound and should not be considered as a tonal component. If the set is accepted as tonal component, the corresponding critical band number is estimated by a weighted average of the critical band numbers $z_t(i)$ in $C_{j_{\max}}$:

$$\bar{z}_{\max} = \frac{\sum_{i\in C_{j_{\max}}} z_t(i) \cdot N'_{\text{tonal}}(i)}{\sum_{i\in C_{j_{\max}}} N'_{\text{tonal}}(i)} \tag{5}$$

The corresponding index for this average band number can be calculated as

$$i_{\max} = \lfloor 2\bar{z}_{\max} \rceil, \tag{6}$$

where the $\lfloor\cdot\rceil$ operator denotes rounding to the nearest integer. This index is used to add the tonal loudness of the tonal component to $N_{\text{tonalComp}}(i)$:

$$N_{\text{tonalComp}}(i) \rightarrow N_{\text{tonalComp}}(i) + \sum_{i\in C_j} N'_{\text{tonal}}(i) \cdot \triangle z, \tag{7}$$

where the $\rightarrow$ operator describes the assignment of the right side to the left side variable. Accordingly, the assignment

$$N_{\text{noiseComp}}(i) \rightarrow N_{\text{noiseComp}}(i) + \sum_{i\in C_j} \left(N'_{\text{signal}}(i) - N'_{\text{tonal}}(i)\right) \cdot \triangle z \tag{8}$$

is made. The next iteration starts by assigning a new set, of indices I without the indices of $C_{j_{\max}}$

$$I \rightarrow I \setminus C_{j_{\max}}. \tag{9}$$

With this set of indices, the iteration starts again with Equation (1). The iterative process is stopped once the condition described in Equation (2) is not fulfilled. The frequency $f_{Comp}(i)$ (corresponding to $\bar{z}_{\max}$) is saved as the frequency of the maximum component assigned to i during all iterations for further analyses (see section 4.2).

3.4 Calculation of Tonality

The mapping of tonal loudness to tonality is done by SNR-dependent scaling of the tonal loudness. The scaling factor $q(i)$ is calculated using the same sigmoid function

$$q(i) = \begin{cases} 1 - \mathrm{e}^{-A\cdot(\mathrm{SNR}(i)-B)}, & \mathrm{e}^{-A\cdot(\mathrm{SNR}(i)-B)} < 1 \\ 0 & \mathrm{e}^{-A\cdot(\mathrm{SNR}(i)-B)} \geq 1 \end{cases} \tag{10}$$

as in ECMA-74. However, the SNR is calculated differently and consequentially the parameters A and B need to be adjusted. The SNR is calculated in dependency of the index i of the critical band:

$$\mathrm{SNR}(i) = \frac{N_{\text{tonalComp}}(i)}{N_{\text{noiseComp}}(i)} \tag{11}$$

and the parameters are set to $A = 3$ and $B = 0.24$. The tonality of each tonal component is then calculated as

$$T_{\text{Comp}}(i) = q(i) \cdot N_{\text{tonalComp}}(i). \tag{12}$$

A calibration factor as described in ECMA-74 is not necessary. $T_{\text{Comp}}(i)$ is actually dependent on the time index l and the critical band index i or the critical band z. For consistency with ECMA-74, it is denoted $T_{\text{Comp}}(l, z)$. The time-dependent tonality $T(l)$ can be calculated by taking the maximum of $T_{\text{Comp}}(l, z)$ over all critical bands z. This calculation procedure is based on the assumption that the perception of tonality is defined by the most dominant tonal component. The single value tonality T is calculated from the time-dependent tonality as explained in ECMA-74.

The unit of the tonality calculated by the Psychoacoustic Tonality method is given in tu_{HMS} (tonality units according to the hearing model of Sottek). The tonality of a 1 kHz tone with a sound pressure level of 40 dB is 1 tu_{HMS} by design of the algorithm.

4. RESULTS

4.1 Effect of revised Outer/Middle Ear Filter

The outer and middle ear filtering was revised because the ECMA-74 tonality often overestimates tonality of low-frequent tonal components. To validate this effect and to design the new filters, listening tests were conducted, which were published in [10]. In these listening tests, the equal loudness contours of the tonal loudness of sinusoids with frequencies below 1 kHz were determined.

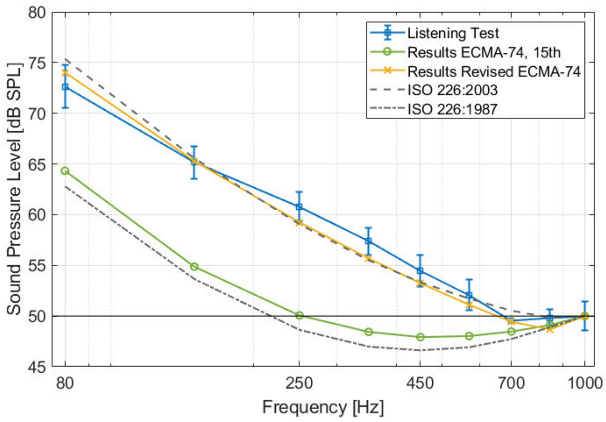

Figure 5 – Equal loudness contour (50 phon) of the tonal loudness for sinusoids below 1 kHz

Figure 5 shows the result for the 50 phon contour. The results of the listening test are close to the ELCs of ISO 226:2003. Since the outer and middle ear filtering in ECMA-74 was designed aiming for the ELCs of ISO 226:1987, the results of ECMA-74 follow this contour closely. This behavior results in the overestimation of the tonality. With the new filter of the revised tonality calculation, the ELCs follow the ISO 226:2003 contour and are thus much closer to the result of the listening test.

4.2 Effect of higher resolution and identification of components

The saved frequency $f_{Comp}(i)$ of the tonal component (see section 3.4) achieved by the revised tonality algorithm provides much clearer results. Figure 6 shows the effect on two synthetic signals.

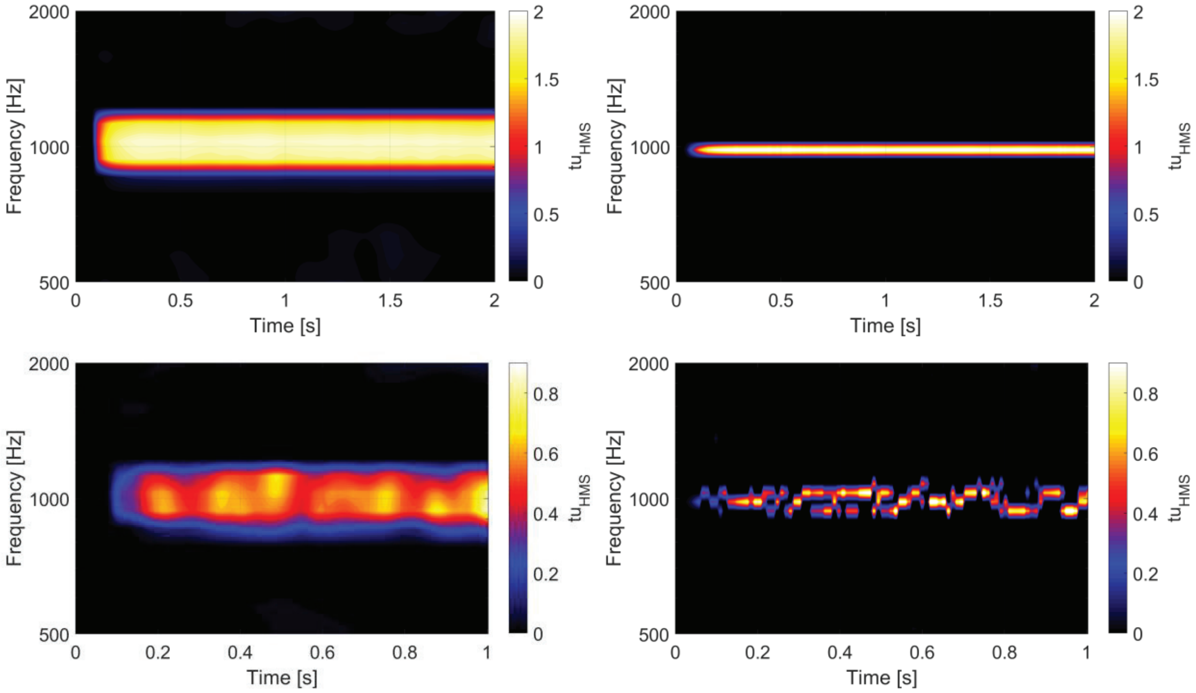

Figure 6 – Left: Results of ECMA-74 tonality, Right: Result of revised tonality algorithm. Top to bottom: Sinusoid (1000 kHz, 65 dB SPL) in pink noise (60 dB SPL), narrowband noise (950-1050 Hz, 65 dB SPL) in pink noise (60 dB SPL)

The result of the sinusoid helps to illustrate how much the resolution is improved. The result of the narrowband noise shows how the improved resolution helps in understanding the character of a tonal sound. The ECMA-74 result could occur from a narrowband noise or from a pure tone with fluctuating frequency and amplitude. With the higher resolution, it is clear that the tonality comes from different frequencies and thus originates from a narrowband noise.

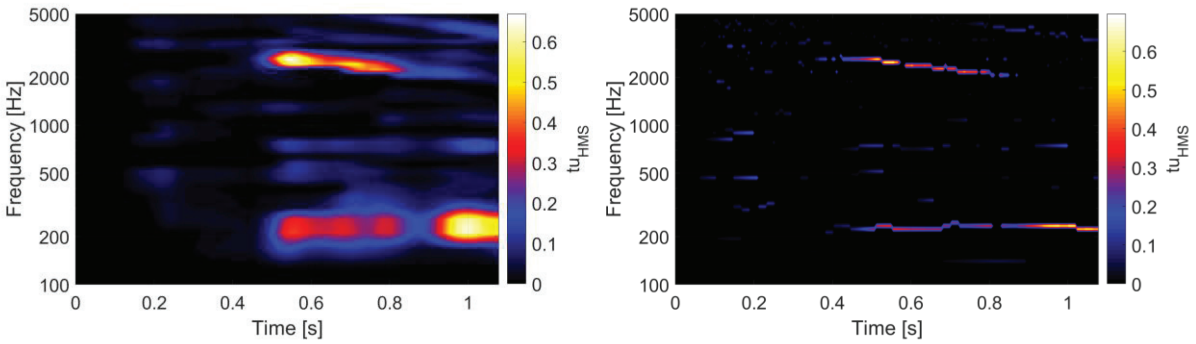

Figure 7 – Engine starter with tonal component. Left: Results of ECMA-74 tonality, Right: Result of revised tonality algorithm.

Figure 7 shows the results of an engine starter which produces a tonal sound. The results show that the amplitude of the tonality is rather consistent for the two calculation methods, while the frequency resolution is improved in the revised algorithm.

5. CONCLUSIONS

In this paper, modifications to the ECMA-74 tonality calculation are presented. Modifications to the outer and middle ear filtering improve the results for low frequencies. Additionally, the new definition of the filter allows different calculations depending on the sound field. The newly introduced grouping of tonal components results in a much higher frequency resolution of the calculated tonalities. Additionally, it inherently suppresses noise and the estimated tonal components can be used for further processing in the future.

The presented modifications will likely be included in a future version of the EMCA-74 standard.

REFERENCES

1. Nobile MA, Bienvenue GR. A Procedure for Determining Prominence Ratio of Discrete Tones in Noise Emissions, NoiseCon 1991, Tarrytown, NY, USA.
2. ECMA-74 15th Edition/July 2018, D.7, Tone-to- Noise Ratio Method; ECMA International, Rue du Rhône 114, CH-1204 Geneva, Switzerland.
3. Hansen H, Verhey JL, Weber R. The Magnitude of Tonal Content. A Review, Acta Acustica united with Acustica, 97(3), pp. 355-363, 2011.
4. Hansen H, Weber R: Zum Verhältnis von Tonhaltigkeit und der partiellen Lautheit der tonalen Komponenten in Rauschen, Proc. DAGA 2010, Berlin, pp. 597-598, 2010.
5. Verhey JL, Stefanowicz S. Binaurale Tonhaltigkeit, Proc. DAGA 2011, Düsseldorf, pp. 827-828, 2011.
6. ECMA-74 15th Edition/July 2018, Annex F and Annex G, ECMA international, Rue du Rhône 114, CH-1204 Geneva, Switzerland.
7. Sottek R. Progress in calculating tonality of technical sounds, Proc. Inter-Noise 2014, Melbourne, 2014.
8. Sottek R. Calculating tonality of IT product sounds using a psychoacoustically-based model, Proc. Inter-Noise 2015, San Francisco, 2015.
9. Pietila G, Seldon W, Roggenkamp T, Bohn T. Tonal Annoyance Metric Development for Automotive Electric Vehicles (No. 2019-01-1467). SAE Technical Paper.
10. Becker J, Sottek R. Tonal Annoyance vs. Tonal Loudness and Tonality, Proc. Inter-Noise 2019, Madrid.
11. Bray W. A New Psychoacoustic Method for Reliable Measurement of Tonalities According to Perception. Proc. Inter-Noise 2018, Chicago, 2018.
12. Sottek R. Modelle zur Signalverarbeitung im menschlichen Gehör, dissertation, RWTH Aachen, 1993.
13. Fastl H, Zwicker E. Psychoacoustics. Facts and models. Springer Verlag, Berlin, Heidelberg, New York, 2006.