Measurement of the attributes of complex tonal components commonly found in product sound

Aaron Hastings,\textsuperscript{a1} Kyoung Hoon Lee,\textsuperscript{a1} Patricia Davies,\textsuperscript{a1} and Aimée M. Surprenant\textsuperscript{b1}

All rotating components in machinery produce sounds that contain tonal components, and the presence of these tones can significantly affect the quality of the product sound. Tone corrections for metrics based on weighted, average sound pressure level have been used since the late 1960s to assess annoyance due to aircraft noise and to rate climate control machinery. Much research has also been focused on measuring the strength of well separated tones in noise. Metrics such as the Prominence Ratio, the Tone-to-Noise Ratio, and variants, as well as more complex models such as the Joint Nordic Method, Aures’ Tonalness, and Virtual Pitch, produce values that often correlate well with subjects’ judgments on the level of the tonal features that they perceive when listening to the sound. However, when sounds are more complex, these metrics do not always work well. Models for the tonalness of two types of tonal sounds are considered here: narrow-band random noise and tones with random frequency fluctuations. The influence of bandwidth and roll-off rate on perceived tonalness are explored for the narrow-band sounds, and the effect of the range and the rate of change of the frequency variation on perceived tonalness is explored for frequency modulated sounds. It was found that roll-off rate affected the perception of tonalness, and that when frequency variations could be tracked or were very small, metrics derived from averaged spectra produced inaccurate predictions of tonalness. Based on the results of these two investigations modifications to tonal metrics are proposed. © 2003 Institute of Noise Control Engineering.

Primary subject classification: 79; Secondary subject classification: 61

1. INTRODUCTION

A. Factors affecting the perception of tonalness

Strong tonal components in a product’s sound can detract from the product’s acceptability. A better understanding of how tonal components are perceived is useful for designers who want to be able to improve the sound quality of their products. Although “tonality” is commonly used to describe the degree to which a sound is tonal, this term is also used in the musical community to describe the organization of all the tones and chords of a piece of music in relation to a tonic.\textsuperscript{1} In order to avoid confusion, the term “tonalness” will be used to describe the degree to which a sound is perceived as tonal. The tonalness of a sound can be evaluated by the sound’s ability to evoke a pitch, i.e. its pitch strength;\textsuperscript{2} by the amount of tonal content, i.e. by the presence of prominent tones;\textsuperscript{3} or by how similar a sound is to a pure tone, i.e. a single sinusoid.\textsuperscript{4} Although there will be differences, for example, a tone in noise can have a strong pitch strength but will not sound like a pure tone, these methods of tonal evaluation should have similar trends for single tonal components that are narrower than a critical band.

Basic factors that affect tonalness include: bandwidth, center frequency, level above threshold, and the number of tonal components present. Wiesmann and Fastl have shown that as the bandwidth increases, the perception of pitch strength decreases for bandpass noise.\textsuperscript{2} Aures modeled the bandwidth effect as being independent of the location of the tonal component’s center on the critical band rate scale.\textsuperscript{5} Aures also found that a frequency of about 700 Hz for pure tones induced the most tonal sensation.\textsuperscript{5} Similar results were obtained by rating pitch strength, with the maximum for pure tones occurring around 1.5 kHz.\textsuperscript{6} Provided that a tonal component is easily perceived, tonalness does not change much due to changes in the absolute level of the tonal component.\textsuperscript{5,7} When additional tonal components are added, the tonalness should change. However, this change will depend on what question is asked during the evaluation. If the question asked is what is the pitch strength or salience, then individual components will tend to be stronger if they are not harmonics of the same fundamental.\textsuperscript{8} If the question asked is how much tonal content the sound has, then for widely spaced components, the overall tonalness will increase equally for harmonic and inharmonic components.\textsuperscript{9} It is presumed that when subjects are asked to evaluate how similar a sound is to a pure tone, additional components will not increase the perception of tonalness.

B. Methods used to predict tonalness and tonal prominence

Existing sound pressure level based noise metrics have been modified to take into account the presence of tones by adding penalties for the presence of tones in sound. These metrics include the 1995 Standard for Sound Rating of Outdoor Unitary Equipment,\textsuperscript{10} the United States Federal Aviation’s regulations for Aircraft Flyovers\textsuperscript{11,12,13} and revised noise criteria for design and rating of heating ventilation and air-conditioning systems.\textsuperscript{14} Although these tonal penalties are useful for determining acceptability for the applications for which they were designed, it is not clear how these penalties

\textsuperscript{a1} The Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907-1077, U.S.A.; E-mail: daviesp@purdue.edu.
\textsuperscript{b1} Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, U.S.A.; E-mail: anm@purdue.edu.
can be extended to other applications because the types of tonal components, other sound characteristics, and the judgments of acceptability can change from application to application.

Developing models relating tonalness to acceptability requires the determination of both tonalness and acceptability. Although the judgment of acceptability must be subjectively determined for each application, tonalness metrics have been developed which can be used to predict tonalness for a wide range of sounds. Being able to calculate tonalness removes the need for one of the time consuming subjective evaluations. Tonalness metrics include the Tone-to-Noise and Prominence Ratio, variants thereof, and Aures model of tonalness. For completeness these metrics are described briefly below.

1. Tone-to-Noise Ratio

The Tone-to-Noise Ratio, \( \Delta L_T \), is defined as the ratio of power contained in the critical band centered on that tone, but not including the part of the spectrum that includes the tonal feature. The tone power, \( W_t \), is determined from the estimated power spectral density of the sound and the width of the tone frequency band, \( \Delta f \) Hz, which is equal to the number of discrete data points included in the bands times the resolution bandwidth. In the standard, the width of the tonal feature is defined by its intersection with the noise floor. The masking noise power, \( W_n \), is calculated by using:

\[
W_n = (W_{tot} - W_t) \frac{\Delta f_c}{(\Delta f_c + \Delta f)},
\]

where \( W_{tot} \) is the total power in the critical band centered on the tone, \( \Delta f_c \) is the width of the critical band and \( \Delta f \) is the width of the frequency band used to compute \( W_{tot} \). \( \Delta f_c \) is close to \( \Delta f \), the differences are caused by the finite resolution of the spectral analysis.

The Tone-to-Noise Ratio, in decibels, is then:

\[
\Delta L_T = 10 \log_{10} \left( \frac{W_t}{W_n} \right) \text{ dB}. \tag{2}
\]

The tone with the highest level in the band is identified as the primary tone, and its frequency is denoted by \( f_p \). For the critical band centered on this primary tone, the tone with the second highest level is identified as the secondary tone and its frequency denoted as \( f_s \). If the secondary tone is sufficiently close in frequency to the primary tone, then the two are considered to be perceived as a single discrete tone and the prominence is determined by combining their tone powers. The proximity criterion is given by the following equation:

\[
f_d = \left| f_s - f_p \right| < 21 \times 10^{1.2 \log_{10} \left( \frac{f_p}{212} \right)^{1.8}} \text{ Hz}, \tag{3}
\]

If the proximity criterion is met, then the power of the secondary tone is added to the power of the primary tone when calculating the tone power \( W_t \). If the proximity criterion is not met, then the tones are considered to be perceived as separate tones and are treated individually. In this case, the power of the secondary tone is removed from the masking noise power before calculating the Tone-to-Noise Ratio of the primary tone. If there are multiple tones within a critical band the same procedure should be repeated for all the tones.

2. Prominence Ratio

The Prominence Ratio, \( \Delta L_p \), is defined as the ratio of the power contained in the critical band centered on the tone under investigation, \( W_{tot} \), to the average power contained in the two adjacent critical bands, one below, \( W_b \), and one above, \( W_u \). The Prominence Ratio, in decibels, is calculated as follows:

\[
\Delta L_p = 10 \log_{10} \left( \frac{W_M}{(W_t + W_{tot})/2} \right) \text{ dB}. \tag{4}
\]

3. Tonal audibility (Joint Nordic Method)

In the Joint Nordic Method, the sound pressure levels of the tones are determined from a narrow-band power spectrum. All local maxima with a 3 dB bandwidth smaller than 10% of the width of the critical band centered on the location of the maximum are regarded as “tones.” This definition includes narrow bands of noise and tones with frequency variation less than 10% of the critical bandwidth. The levels, \( L_p \) of all tones in the same critical band are added to give the total tone level for the band, \( L_p \):

\[
L_p = 10 \log_{10} \sum 10^\frac{L_p}{10} \text{ dB}. \tag{5}
\]

The total sound pressure level of the masking noise, \( L_{pm} \) in a critical band is determined from the average noise level within that critical band, \( L_{pn,avg} \) which may be found by visual averaging of the narrow-band frequency spectrum while disregarding all maxima in the spectrum resulting from tones and their possible side bands in that range. Thus:

\[
L_{pm} = L_{pn,avg} + 10 \log_{10} (\text{number of spectral lines in the critical band}). \tag{6}
\]

The Tonal Audibility, \( \Delta L_a \), is expressed in dB above the masking threshold, \( MT \) by:

\[
\Delta L_a = L_p - L_{pm} + 2 + 10 \log_{10} \left( \frac{1 + \left( \frac{f_p}{502} \right)^{25}}{0.5} \right) \text{ dB, ref. MT}. \tag{7}
\]

In the Joint Nordic Method this Tonal Audibility is used to penalize the measured average A-weighted sound pressure level, \( L_{eq,avg} \), so that the resulting corrected sound pressure level can be used as a predictor of the impact of the sound. The penalty, \( k \), is:

\[
k = \begin{cases} 
0 \text{ dB} & \text{for } \Delta L_a < 4 \text{ dB}, \\
\Delta L_a - 4 & \text{for } 4 \text{ dB} \leq \Delta L_a \leq 10 \text{ dB}, \\
6 \text{ dB} & \text{for } \Delta L_a > 10 \text{ dB}.
\end{cases} \tag{8}
\]

4. Aures’ model of tonalness

Aures modeled tonalness based on subjective evaluation of pure tones and bandpassed noise by using three weighting functions for the effects of bandwidth, center frequency, and prominence. The first of these weighting functions is:

\[
w_t(\Delta f) = \frac{0.13 + \Delta f}{0.13 + \Delta f}, \tag{9}
\]

where \( \Delta f \) is the bandwidth of the tonal component as a fraction
of the critical bandwidth. The frequency weighting function is:

$$w_2(f) = \left(1 + 0.2 \left(\frac{f}{700} + \frac{700}{f}\right)^2\right)^{-0.29},$$

(10)

where \(f\) is the center frequency of the tonal component in Hz, and the prominence weighting function is:

$$w_3(\Delta L) = \left(1 - e^{-\Delta L/15}\right)^{0.29},$$

(11)

where \(\Delta L\) is the excess level in dB defined by:

$$\Delta L = L_i - 10 \log_{10} \left[\sum_{k} A_{k,i} (f_i)\right]^2 + E_G(f_i) + E_{HS}(f_i),$$

(12)

and \(L_i\) is the power of the tonal component of interest in dB, \(A_{k,i}\) is the secondary excitation at \(f_i\) due to the \(k_{th}\) component, \(E_G\) is the masking intensity of the noise, and \(E_{HS}\) is the intensity at the threshold of hearing. Aures combines these into \(w_r\):

$$w_r = \sum_{k} \left[w_1(\Delta L_i) w_2(f_i) w_3(\Delta L_i)\right]^2,$$

(13)

where \(w_1 = w_1^{1.029}, w_2 = w_2^{1.029},\) and \(w_3 = w_3^{1.029}\). A weighting factor based on loudness of the noise component, \(N_{Gr}\), and the loudness of the overall sound, \(N_{Total}\) is also calculated:

$$w_{Gr} = 1 - \frac{N_{Gr}}{N_{Total}}$$

(14)

This is the fraction of the total loudness caused by the presence of the tonal components. Finally, Aures Tonalness is:

$$T = c \cdot w_r^{0.29} w_{Gr}^{0.79},$$

(15)

where \(c\) is a constant chosen such that a 1 kHz pure tone with a level of 60 dB would have a tonalness of 1. For an ideal implementation of this model, \(c\) should be approximately 1.09.

C. Observed deficiencies in the models

The Aures model predicts tonalness, whereas the other metrics predict the prominence of the tone, which is not necessarily a good predictor of tonalness. Shown in Fig. 1(a) is a plot of Aures Tonalness and the Tone-to-Noise Ratio as a function of the signal to noise ratio in the critical band centered about the tone for a series of pure tones in broad-band noise. For this type of sound, all metrics should work well. While the Tone-to-Noise Ratio increases with increasing signal to noise ratio, the Aures Tonalness saturates after a certain level of tonal prominence has been reached. Tone-to-Noise Ratio, Prominence Ratio, and Tonal Audibility, \(\Delta L_{TA}\), are more closely related to the \(\Delta L\) term defined in Eq. (12), as can be seen in Fig. 1(b). Recognition of the presence of a saturation effect results in the limitation of the tone penalty to a maximum of 6 dB in the Joint Nordic Method. However, the results here show that the perception of tonalness, as predicted by Aures model continues to increase well beyond a Tonal Audibility level of 10 dB. However, it may well be that the increase in tonalness beyond the level of 10 dB may not strongly affect the annoyance level.

The measurement of bandwidth and isolation of the “tone power” from the “noise power” is problematic with most of these methods. As will be shown later, some sounds are perceived to be tonal, yet use of the conventional 3 dB down or intersection with the noise floor approaches to measuring bandwidth results in predictions of tonal prominence or tonalness that are too low. Even wide-band filtered random noise can sound tonal and the sensation can increase if the filter roll-off rate is increased.\(^{17}\) Analysis of frequency modulated tones in noise is also problematic, as the averaging required to produce low variance spectral estimates results in tonal features in the spectrum being too wide. When the variation in the frequency can be tracked the tonal feature may sound almost as tonal as a stationary tone. The answer should be shorter analysis times which should result in narrower spectral features, but this leads to higher variability in spectral estimates and/or poor spectral resolution, both of which lead to problems when using the spectra to calculate the metrics described above.

Two sets of studies related to the perception of tonalness and its relationship to tonal prominence are described below. In the first study the effect of roll-off rate on the perception of
the tonalness of bandpass filtered noise was examined. The second study was focused on developing an understanding of how frequency modulation affects tonalness. Based on the results of both studies, suggestions for improving the metrics described above are given, as well as suggestions for further research.

2. TONALNESS OF NARROW-BAND RANDOM NOISE

A pure tone can be thought of as the archetype for tonal stimuli. A pure tone excites the narrowest region along the basilar membrane; is the most stationary type of sound; and induces the strongest sensation of pitch. In contrast, random uniformly exciting noise can be thought of as the archetype for non-tonal stimuli. It excites the entire basilar membrane; is constantly and randomly changing level at all frequencies; and evokes no sensation of pitch. However, if this noise is filtered by a narrow bandpass filter, it takes on some of the characteristics of a pure tone. As the filter becomes narrower, the region excited by bandpassed noise becomes smaller and the random changes become smaller. If an infinitely narrow bandpass filter were to be used to filter the noise, the signal would become a pure tone. Thus it is reasonable to expect that there should be a continuous transition from the non-tonalness of noise to the tonalness of a pure tone and that this transition should be dependent on the narrowness of the bandpass filter used to filter the noise.

The bandwidth of bandpassed noise is often described by the frequency width of the filter’s passband. For example, Butterworth bandpass filters have cut-off frequencies defined by the frequencies at which the power has dropped to one-half the peak power (3 dB down point), as shown in Fig. 2(a). Tonalness experiments using bandpassed noise conducted by Aures, Wiesmann and Fastl, and Vormann et al. have all shown that as bandwidth increases, tonalness decreases. However, for the same bandwidth, one would expect filters with lower roll-off rates to be perceived as less tonal than filters with higher roll-off rates for three reasons. Because low roll-off rate filters are less effective at attenuating noise far from the cut-off frequencies, there is more noise in the sound which widens the excitation region along the basilar membrane and increases the random character of the sound. Because much of this noise is close in frequency to the component of the sound that is perceived as tonal, this increased noise can also partially mask the tonal character. Finally, high roll-off rates can evoke a tonal perception. It is hypothesized that a model that accounts for both the peak bandwidth and the roll-off rate will predict the tonalness of bandpass noise better than Aures’ bandwidth weighting function, w_p.

A. Tonalness experiments

Three sets of psychoacoustic experiments are described below. The experiments were performed to verify Aures’ model, to determine whether the psychoacoustic testing method affected the tonalness evaluations, and to examine how roll-off rate affected the tonalness of the sound.

1. Tonalness of pure tones and bandpass filtered noise

In a preliminary experiment digital Butterworth bandpass filters were used to filter Gaussian distributed white noise to produce a set of stimuli for subjects to rate in terms of tonalness. The test was conducted to confirm the bandwidth weighting function, w_p, and the frequency weighting function, w_f, in the model of tonalness proposed by Aures. The parameters used to generate these sounds are shown in Table 1. The bandwidth of the Butterworth filters were defined by the 3 dB down points and were then related to the critical bandwidth as given in Table 6.1 of Psychoacoustics, Facts and Models for the center frequencies shown in Table 1. The tonalness for these test sounds was determined by using Aures’ Tonalness model and from subjective evaluations in paired comparison tests. The Bradley-Terry-Luce model was used to transform the response data to tonalness ratings. Each group of sounds described in Table 1: sounds 1-8, 9-16, 17-24, 25-31, 32-38, and 39-45 were used as the stimuli in six paired comparison tests.

In the paired comparison tests, subjects sat in a quiet room designed for subjective testing in front of a computer keyboard and monitor. The computer, which was outside of the room, was connected to the monitor and keyboard via a sealed break.
TABLE 1—Center frequencies and bandwidths of sounds used in preliminary experiment. * indicates a pure tone. ** indicates 6th order Butterworth filters. All other sounds are generated by using 4th order Butterworth filters.

<table>
<thead>
<tr>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5</td>
<td>0%</td>
<td>9</td>
<td>4.5</td>
<td>25%</td>
<td>17</td>
<td>13.5</td>
<td>10%</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>0%</td>
<td>10</td>
<td>5.5</td>
<td>25%</td>
<td>18</td>
<td>13.5</td>
<td>25%</td>
</tr>
<tr>
<td>3</td>
<td>9.5</td>
<td>0%</td>
<td>11</td>
<td>9.5</td>
<td>25%</td>
<td>19</td>
<td>13.5</td>
<td>40%</td>
</tr>
<tr>
<td>4</td>
<td>10.5</td>
<td>0%</td>
<td>12</td>
<td>10.5</td>
<td>25%</td>
<td>20</td>
<td>13.5</td>
<td>70%</td>
</tr>
<tr>
<td>5</td>
<td>12.5</td>
<td>0%</td>
<td>13</td>
<td>12.5</td>
<td>25%</td>
<td>21</td>
<td>13.5</td>
<td><strong>10%</strong></td>
</tr>
<tr>
<td>6</td>
<td>13.5</td>
<td>0%</td>
<td>14</td>
<td>13.5</td>
<td>25%</td>
<td>22</td>
<td>13.5</td>
<td><strong>25%</strong></td>
</tr>
<tr>
<td>7</td>
<td>16.5</td>
<td>0%</td>
<td>15</td>
<td>16.5</td>
<td>25%</td>
<td>23</td>
<td>13.5</td>
<td><strong>40%</strong></td>
</tr>
<tr>
<td>8</td>
<td>17.5</td>
<td>0%</td>
<td>16</td>
<td>17.5</td>
<td>25%</td>
<td>24</td>
<td>13.5</td>
<td><strong>70%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
<th>Sound</th>
<th>f_c</th>
<th>BW</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>4.5</td>
<td>0%</td>
<td>32</td>
<td>9.5</td>
<td>0%</td>
<td>39</td>
<td>12.5</td>
<td>0%</td>
</tr>
<tr>
<td>26</td>
<td>4.5</td>
<td>5%</td>
<td>33</td>
<td>9.5</td>
<td>5%</td>
<td>40</td>
<td>12.5</td>
<td>5%</td>
</tr>
<tr>
<td>27</td>
<td>4.5</td>
<td>10%</td>
<td>34</td>
<td>9.5</td>
<td>10%</td>
<td>41</td>
<td>12.5</td>
<td>10%</td>
</tr>
<tr>
<td>28</td>
<td>4.5</td>
<td>20%</td>
<td>35</td>
<td>9.5</td>
<td>20%</td>
<td>42</td>
<td>12.5</td>
<td>20%</td>
</tr>
<tr>
<td>29</td>
<td>4.5</td>
<td>25%</td>
<td>36</td>
<td>9.5</td>
<td>25%</td>
<td>43</td>
<td>12.5</td>
<td>25%</td>
</tr>
<tr>
<td>30</td>
<td>4.5</td>
<td>30%</td>
<td>37</td>
<td>9.5</td>
<td>30%</td>
<td>44</td>
<td>12.5</td>
<td>30%</td>
</tr>
<tr>
<td>31</td>
<td>4.5</td>
<td>35%</td>
<td>38</td>
<td>9.5</td>
<td>35%</td>
<td>45</td>
<td>12.5</td>
<td>35%</td>
</tr>
</tbody>
</table>

The sounds were stored as 16 bit .WAV files on the computer and played out through a LynxONE audio card. A Tucker-Davis Technologies HB7 headphone driver was used to power Etymotic Research ER-2 earphones, which were used to present the sounds to the subjects. The playback system was calibrated prior to each subject participating in the experiment. This hardware setup was used for all subjective tests described in this paper. In all tests, the subjects read and signed a consent form approved by Purdue University’s Committee on the Use of Human Research Subjects and were given a hearing test to make sure that they had normal hearing (no more than 20 dB above the threshold for normal hearing) in the frequency range between 125 and 8000 Hz. Subjects who passed the hearing test had sound samples played to them which included bandpassed noise, pure tones, and white noise and were given a practice test in order to familiarize themselves with making tonal judgments. During the test, the subjects were asked to “choose the sound that was more tonal”.

In order to compare both the Aures and paired comparison results, the tonalness values obtained from the Bradley-Terry-Luce model were rescaled by using a first order linear regression to minimize the error, $\varepsilon_{ij}$, in $y_{ij} = \beta_j + \alpha x_{ij} + \varepsilon_{ij}$, where $y$ is the calculated tonalness from Aures’ Tonalness model, $x$ is the tonalness values estimated using the Bradley-Terry-Luce model on the subjective test data, $i$ is the sound number within a subjective test, $j$ is test number, and $\alpha$ and $\beta$ are constants to be determined. Thus each test had a unique $\beta$ but all tests shared the same $\Delta$. This rescaling assumed that subjects did not change their judgment criteria between tests. Provided that the subjects were able to make judgments on some type of gestalt tonalness this assumption should be valid. Because different tests varied either by center frequency, filter order, or bandwidth, it is possible that subjects did change their judgment criteria, however, as a preliminary experiment, it was assumed that this rescaling would be sufficient to provide guidance for designing more detailed experiments.

Aures’ model predictions and the rescaled Bradley-Terry-Luce tonalness values are shown in Fig. 3. The results support the frequency weighting function given by Aures for both pure tones and bandpassed noise when the bandwidth was 25% of the critical bandwidth centered at $f_c$. When bandwidth is increased tonalness decreases both in Aures’ model and in the subjective test results. However, test sounds with the same 3 dB down bandwidths but different roll-off rates received different tonalness ratings. In addition, the rescaled subjective results were sometimes negative and while the ordering of the ratings was consistent, the model and the subjective test results did not always align. The discrepancies are more clearly illustrated in Fig. 4. The differences may be the result of the different methods used to produce different bandwidths as a percent of the critical bandwidth. Aures generated critical bandwidth variations from 2% to 25% by varying the center frequency of a filter with a frequency bandwidth, $\Delta f$, of 30 Hz. In the present experiment, the center frequency was held constant and $\Delta f$ was varied in order to generate the critical bandwidth variation.

2. Comparison of Direct Scaling and Paired Comparison Testing

The differences found in the bandwidth effect, between the subjective results and Aures’ model are of interest since for a single bandpassed component, the model’s greatest variation is due to the bandwidth effect, thus errors in the bandwidth weighting function, $w_i$, can lead to significant errors in the estimated tonalness. In order to have finer control over the roll-off rate and peak bandwidth, an experiment where trapezoidal filters were used to generate sounds covering a wide tonal range was designed. The peak bandwidth of a trapezoidal

Fig. 3—Tonalness calculated from Aures’ model ( - - open symbol) and from 6 paired comparison experiments rescaled by using a linear regression ( – solid symbol). The parameters used to generate each sound are given in Table 1.
filter is defined by the frequencies where the magnitude of the frequency response of the filter just begin to roll-off (where the gain becomes less than 0 dB), as shown in Fig. 2(b). Filter peak bandwidths were chosen such that they would be a percent of the critical bandwidth (CBW) where:

\[
\text{CBW} = 24.7\left(\frac{f_c/1000+1}{f_c}\right),
\]

and \(f_c\) is the center frequency of the filter in Hz.\(^{21}\) To achieve the desired roll-off characteristics, zero-phase finite impulse response (FIR) filters were designed by specifying the frequency response magnitude at intervals of 2.5 Hz. The sample rate was 22,050 samples per second resulting in 8820 point filter. The frequency responses of the resulting filters were checked to ensure that the filters were close to the desired roll-off at frequencies in between those specified by the design. An example showing the detail around the lower cut-off frequency is shown in Fig. 5.

These trapezoidal filters were used to filter 20 seconds of random noise. The sounds were normalized to 16 sones for a diffuse sound field according to ISO 532B.\(^{22}\) During subjective testing, only the middle 2.5 seconds of each sound were used. This ensured that there was no ringing effect present in the stimuli due to the transient response of the FIR filters. To avoid “clicks” during testing, twenty milli-second cosine ramps were applied to the beginning and end of each sound.

In the first experiment described above, six paired comparison tests were performed to determine the tonalness of the 45 sounds, however, these tonalness values were not fixed to an absolute point on a common tonalness scale, so Aures’ model was used. To determine the relative tonalness of all 45 sounds without using such additional information, all sounds would have to be evaluated within the same paired comparison test. However, the paired comparison method is impractical for tests with more than 12 sounds because the tests take too long and the subjects begin having problems concentrating on the task. Therefore, the next experiment was designed to determine if reliable results could be obtained by using direct scaling. This was done by subjectively evaluating a large group of sounds by using direct scaling and a subset of these sounds by using paired comparisons. If the subjective tonalness results were in agreement for both methods, then a larger experiment using only direct scaling could be designed.

Forty-six bandpassed random noise signals, white noise, and a pure tone were evaluated by 22 subjects in a direct scaling test and 12 of the bandpassed signals were re-evaluated in a paired comparison test by the same subjects. The pure tone and all the bandpassed sounds were centered at 700 Hz. The peak bandwidth and roll-off rates for these tests are shown in Table 2. The sounds used in the direct scaling are indicated by “+” signs and the sounds used in the paired comparison test are indicated by “#” signs.

Prior to the test, samples of white noise, a pure tone, and 16 bandpassed noise signals covering the tonal range were then presented to them. The subjects were encouraged to play these sample sounds as often as they felt it was necessary. The subjects were told that the white noise should be considered as not at all tonal and the pure tone should be considered as purely tonal. For the direct scaling test the subjects were told that they should give a rating of 2 to a sound that was not at all tonal and a value of 8 for a pure tone, however, they were allowed to go beyond this scale if they felt it was necessary, for example, if they had rated one sound as an 8 but then later heard a more tonal sound, they could give the second sound a value greater than 8. In the paired comparison test, the subjects were asked to choose the sound in each pair that was more tonal. Ties were not allowed. The paired comparison results were used to estimate the tonalness by using the Bradley-
Influences of center frequency, bandwidth, and roll-off rate

In the final filtered noise experiment, the same trapezoidal filter design described in the previous section was used to generate bandpassed noise. The sounds generated had peak bandwidths ranging from 1 to 200% of the critical bandwidth and attenuation rates ranging from 20 to 300 dB per octave. Because the auditory filter changes shape with level and center frequency, tests were conducted at 16 sones for sounds centered at 300, 500, 700, 2000, and 4000 Hz and additionally at 8 sones for sounds centered at 700 Hz.

Six tests were conducted, one for each center frequency and loudness combination. The signals in each test consisted of a pure tone, white noise, and 38 bandpassed sounds generated by using trapezoidal filters whose roll-off rates and peak bandwidths are shown in Table 3. The 7 male and 10 female subjects were between 23 and 37 years of age and were from a number of different countries: China, Columbia, Germany, Greece, India, Korea, Turkey and the United States of America. All 17 subjects had normal hearing. The subjects were asked to evaluate the tonalness of the stimuli. The subjects took two practice tests, one with sounds centered at 500 Hz and one centered at 2000 Hz in order to familiarize themselves with making judgments on the given scale. The subjects were told that they could repeat the practice tests until they felt comfortable making the judgments. In the tests, the order of the stimuli was randomized and a different order was used each time the test was run.

Multiple runs were conducted for two subjects in order to examine repeatability. The within subject variation for 5 runs of the test at 700 Hz and 16 sones is shown for subject A (high repeatability) in Fig. 7(a) and for subject B (low repeatability) in Fig. 7(b). The order of the sounds was randomized and different for each subject and each of the 5 runs of the test, therefore some of the variability observed is due to ordering effects. The ratings for sounds with average scores between 3 and 6 had greater variation than the ratings for sounds that had low or high tonalness. Multiple runs for the same subject were also used to determine if the subjects were still adjusting their criteria for judging tonalness. If there was a consistent

<table>
<thead>
<tr>
<th>Roll-Off Rate, dB per Octave</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBW</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

3. Influences of center frequency, bandwidth, and roll-off rate

In the final filtered noise experiment, the same trapezoidal filter design described in the previous section was used to generate bandpassed noise. The sounds generated had peak bandwidths ranging from 1 to 200% of the critical bandwidth and attenuation rates ranging from 20 to 300 dB per octave. Because the auditory filter changes shape with level and center frequency, tests were conducted at 16 sones for sounds centered at 300, 500, 700, 2000, and 4000 Hz and additionally at 8 sones for sounds centered at 700 Hz.

Six tests were conducted, one for each center frequency and loudness combination. The signals in each test consisted of a pure tone, white noise, and 38 bandpassed sounds generated by using trapezoidal filters whose roll-off rates and peak bandwidths are shown in Table 3. The 7 male and 10 female subjects were between 23 and 37 years of age and were from a number of different countries: China, Columbia, Germany, Greece, India, Korea, Turkey and the United States of America. All 17 subjects had normal hearing. The subjects were asked to evaluate the tonalness of the stimuli. The subjects took two practice tests, one with sounds centered at 500 Hz and one centered at 2000 Hz in order to familiarize themselves with making judgments on the given scale. The subjects were told that they could repeat the practice tests until they felt comfortable making the judgments. In the tests, the order of the stimuli was randomized and a different order was used each time the test was run.

Multiple runs were conducted for two subjects in order to examine repeatability. The within subject variation for 5 runs of the test at 700 Hz and 16 sones is shown for subject A (high repeatability) in Fig. 7(a) and for subject B (low repeatability) in Fig. 7(b). The order of the sounds was randomized and different for each subject and each of the 5 runs of the test, therefore some of the variability observed is due to ordering effects. The ratings for sounds with average scores between 3 and 6 had greater variation than the ratings for sounds that had low or high tonalness. Multiple runs for the same subject were also used to determine if the subjects were still adjusting their criteria for judging tonalness. If there was a consistent

Table 2: Peak bandwidths and roll-off rates used to generate bandpassed noise signals: "+" direct scaling test stimulus, "#" paired comparison test stimulus.

<table>
<thead>
<tr>
<th>Roll-Off Rate, dB per Octave</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBW</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

3. Influences of center frequency, bandwidth, and roll-off rate

In the final filtered noise experiment, the same trapezoidal filter design described in the previous section was used to generate bandpassed noise. The sounds generated had peak bandwidths ranging from 1 to 200% of the critical bandwidth and attenuation rates ranging from 20 to 300 dB per octave. Because the auditory filter changes shape with level and center frequency, tests were conducted at 16 sones for sounds centered at 300, 500, 700, 2000, and 4000 Hz and additionally at 8 sones for sounds centered at 700 Hz.

Six tests were conducted, one for each center frequency and loudness combination. The signals in each test consisted of a pure tone, white noise, and 38 bandpassed sounds generated by using trapezoidal filters whose roll-off rates and peak bandwidths are shown in Table 3. The 7 male and 10 female subjects were between 23 and 37 years of age and were from a number of different countries: China, Columbia, Germany, Greece, India, Korea, Turkey and the United States of America. All 17 subjects had normal hearing. The subjects were asked to evaluate the tonalness of the stimuli. The subjects took two practice tests, one with sounds centered at 500 Hz and one centered at 2000 Hz in order to familiarize themselves with making judgments on the given scale. The subjects were told that they could repeat the practice tests until they felt comfortable making the judgments. In the tests, the order of the stimuli was randomized and a different order was used each time the test was run.

Multiple runs were conducted for two subjects in order to examine repeatability. The within subject variation for 5 runs of the test at 700 Hz and 16 sones is shown for subject A (high repeatability) in Fig. 7(a) and for subject B (low repeatability) in Fig. 7(b). The order of the sounds was randomized and different for each subject and each of the 5 runs of the test, therefore some of the variability observed is due to ordering effects. The ratings for sounds with average scores between 3 and 6 had greater variation than the ratings for sounds that had low or high tonalness. Multiple runs for the same subject were also used to determine if the subjects were still adjusting their criteria for judging tonalness. If there was a consistent

Table 3: Peak bandwidths and roll-off rates used to design the trapezoidal filters that were used to generate the test sounds at each center frequency.

<table>
<thead>
<tr>
<th>Roll-Off Rate, dB per Octave</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBW</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>
Modification of Aures’ bandwidth model

earlier in the paper, indicated that the tonalness does vary with frequency and that Aures’ model of this effect was accurate for pure tones. Therefore to combine the results from different tests, this frequency effect needs to be incorporated into the scaling; this was done in the following manner. The pure tone results were set to the result predicted by Aures’ model and the white noise results were set to have a tonalness of 0. The scores resulting from this transformation will be referred to as the re-scaled scores.

In Fig. 9, the tonalness scores are plotted against the filter width expressed as a percentage of critical bandwidth for two roll-off rates: 100 and 300 dB per octave. Two approaches to calculating bandwidth were explored, peak bandwidth, as described earlier, and equivalent rectangular bandwidth (ERB). The higher tonalness ratings for the sounds with the higher roll-off rate is clearly illustrated in this figure, as is the decay in perceived tonalness as the filter bandwidth increases. The differences in the perceived tonalness are also dependant on the center frequency of the band; the differences between the results for the two roll-off rates gets slightly larger as the center frequency of the band increases. This effect was preserved even when the ERB method was used to calculate the bandwidth.

Aures found that bandpassed sounds with bandwidths 140% greater than a critical band had less than 10% of the tonalness of a pure tone. In contrast, in these tests bandpassed noise with a peak bandwidth of 200% of a critical bandwidth and roll-off rate of 300 dB per octave was judged, on average, to have 50% of the tonalness of a pure tone, indicating that roll-off rate as well as bandwidth affects the perception of tonalness. For narrow peak bandwidths, roll-off rate continues to contribute to the tonalness up to at least 300 dB per octave, however, for broader peak bandwidths, increases in the roll-off rate beyond 100 dB per octave become less significant. This saturation effect is less significant at higher center frequencies. Zwicker and Fastl found saturation of tonalness at higher roll-off rates when studying low-pass filtered noise. They found that the roll-off rate saturated at around 24 dB per octave for a cut-off frequency of 250 Hz and at 36 dB per octave for a cut-off frequency of 1000 Hz. For low-pass filtered noise with a 250 (or 1000) Hz cut-off frequency, increasing the roll-off rates of the filters above 24 (or 36) dB per octave does not change the masking pattern, indicating that the auditory filter is responsible for the saturation. The saturation effects observed in the experiments described in this paper occurred at much higher roll-off rates. The lower frequency slope of the auditory filters are much steeper than the upper frequency slope, so it can be argued that the roll-off rate needs to be much higher before the lower frequency masking pattern stops changing. However, it is also possible that tonal perceptions from both the high- and low-frequency cut-offs may be interacting.

**B. Modification of Aures’ bandwidth model**

It has been shown that both the roll-off rate and peak bandwidth are important for the prediction of the tonalness of bandpassed noise and that center frequency plays a role beyond that accounted for in Aures’ $w_f$ model. To improve Aures’ model, the approach adopted here is to modify Aures’ $w_f$ model.
Fig. 8–Average score (17 subjects) for sounds centered at different center frequencies: (a) 300 Hz, (b) 500 Hz, (c) 700 Hz at 16 sones, (d) 700 Hz at 8 sones, (e) 2000 Hz, and (f) 4000 Hz. (−−−) pure tone, (−−) white noise, (−o) bandpassed noise.
Fig. 9—Bandwidth effect for filtered noise with a roll-off rate of 100 dB per octave (–Δ–) and 300 dB per octave (–∇–), for peak bandwidths centered at: (a) 300 Hz, (b) 700 Hz, and (c) 4000 Hz; and for equivalent rectangular bandwidths centered at: (d) 300 Hz, (e) 700 Hz, and (f) 4000 Hz. Solid vertical lines represent the standard deviation of the estimated mean.
model to include the roll-off rate saturation behavior observed in the experiments. Based on observations of the tonalness versus roll-off rate, a model of the form:

\[ w_{\text{mod}} = \left[ \frac{a}{a + \Delta z} \right] \cdot \left[ 1 - \exp\left( \frac{-\text{ROR}}{c + d \cdot z_c} \right) \right]^b, \]  

(15)

where \( \Delta z \) is the sound's equivalent rectangular bandwidth as a fraction of the critical bandwidth, ROR is the roll-off rate in dB per octave, and \( z_c \) is the center frequency of the tonal feature in Bark. \( a, b, c, \) and \( d \) are constants, determined by doing a nonlinear least squares fit of \( w_{\text{mod}} \cdot w_z \cdot w_j \cdot w_{Gr}^{0.79} \) to the re-scaled data from the third set of tests. The first term in Eq. (15) has the same structure as the \( w_1 \) term in Aures' model. The \( c + d \cdot z_c \) term can be thought of as a frequency dependant roll-off-rate constant. As \( z_c \) increases the constant becomes larger and saturation occurs at a higher roll-off rate. Saturation will occur above \( 3 (c + d \cdot z_c) \) after which the relative bandwidth effect will dominate.

For the purpose of determining the noise and tonal content in the sounds for calculating \( w_2, \) \( w_3, \) and \( w_{Gr}, \) the tonal region was estimated by the intersection with the noise floor. Typically, if this region exceeds the critical band, then the tonal region is limited by the critical band as shown by region A in Fig. 10. However since these sounds had very simple shapes, this left a region of sound that, if considered as noise, could effect the tonalness calculation. In order to get the best fit possible, the tonal region was calculated twice, once by using region A and once by using region A+B. The correlation to the subjective results for both calculations are shown in Fig. 11. Fitting this model to the data yielded \( a = 2.02, \) \( b = 1.1, \) \( c = 70, \) and \( d = 0.57 \) for Eq. (15). As can be seen in Fig. 11, reasonably good agreement is obtained.

![Image](image.png)

**Fig. 10**–Illustration of two possible regions for summing the tonal energy. (A) The critical band limits the region of tonal summation. (A+B) The critical band does not limit the region of tonal summation.

**Fig. 11**–Correlation between subjective tonalness using \( w_{\text{mod}}, \) \( w_z, \) \( w_j, \) and \( w_{Gr}^{0.79} \) when (a) the energy due to the tonal component is restricted to be within a single critical band, (b) the energy due to the tonal component is allowed to exceed a single critical band.

### 3. TONALNESS OF FREQUENCY MODULATED SOUNDS

When there are multiple tones in a single critical band, the actual perception of the tone width is influenced by many things like the relative amplitude of tones and the distance between the tones. If there are enough discrete spectral components in a certain frequency band, it is hard to distinguish the discrete components from bandpassed random noise with the same frequency band, as Hartmann, McAdams, Gerzso, and Boulez found in experiments using multiple sine waves. When the frequency modulations change very slowly and are trackable, the sound may sound as tonal as a stationary tone, but by changing the frequency content and range of the random fluctuations, it is possible to generate sounds of significantly different character. The bandwidth of the tonal feature that is estimated from a spectrum may not be an appropriate measure of the “thinness” of the tonal

---

component, because the windowing and averaging used in the spectral estimation may lead to features that are broad in frequency. Since all the metrics described earlier rely on accurate spectral estimates, analysis of these nonstationary sounds is problematic. The temporal windowing in the spectral estimates has to somehow reflect the ability of the human hearing system to track frequency variations in the sound. Too broad a window will lead to spectral smearing with these FM sounds, and too short of a window will lead to too much energy leakage, again producing too broad of a spectral feature. Following is a description of a psychoacoustic test in which subjects were asked to evaluate the tonalness of tones with randomly varying frequency. The performance of the metrics, described above, as predictors of tonalness or tonal prominence for these types of FM signals is also discussed.

A. Frequency modulated test signals

The sounds used in the psychoacoustic tests were frequency modulated tones of the form:

$$y(t) = A\sin\left(2\pi f_c t + 2\pi B \int r(t)\,dt\right),$$

where $r(t)$ was uniformly distributed random noise passed through a fourth order Butterworth low-pass filter with a cut-off frequency $f_c$ and scaled to have a maximum deviation of 1 Hz. The parameter $B$ controls the range of the frequency modulation and the instantaneous frequency of the signal is:

$$f(t) = f_c + Br(t) \text{ Hz}.$$  

(17)

The center frequency, $f_c$, was 700 Hz for all the sounds used in the tests. Sixteen sounds with all combinations of modulation ranges: $B = 25, 50, 75$ and 100 Hz, and filter cut-off frequencies: $f_o = 10, 50, 100$ and 200 Hz were generated at a sample frequency of 44.1 kHz. The power spectra of the signals are shown in Fig. 12 and a 0.5 second snapshot of the instantaneous frequency of each of the signals is shown in Fig. 13. All sounds were 3 seconds long, and 20 ms cosine ramps were applied to the start and end of the signals to avoid clicks during playback. The sounds were chosen so that each had a perceptually different character. Modulation ranges and filter cut-off frequencies were varied until the character of the sound was deemed to have changed noticeably.

As the frequency content of the modulation becomes richer, (moving from left to right in Fig. 12) there is a broadening of the spectrum, but close to the center frequency the spectral feature becomes narrower, and the spectrum resembles that of a tone in noise. As the modulation range ($B$) becomes larger (moving top to bottom in Fig. 12), the spectral feature broadens, as expected.

B. Subjective tests with FM sounds

The sixteen sounds were grouped into three sets of eight sounds and a paired comparison test was used to evaluate each set of eight sounds. The sounds with the two modulation ranges $B = 25$ and $B = 50$ Hz made up set 1; the sounds where $B = 75$ and $B = 100$ Hz made up set 2; and set 3 was composed of sounds with the filter cut-off frequency at $f_{\text{cut}} = 50$ and at 200 Hz. Six subjects evaluated the tonalness of the sounds by choosing the sound that was most like a pure tone. All subjects had normal hearing.

The results of the paired comparison tests were analyzed by using the Bradley-Terry-Luce model.20 The mean values of the scores for each of the tests were adjusted so that signals that appeared in two of the three tests were rated similarly. While it is risky to compare results from different paired comparison tests as judgment criteria may change from test to test, it was felt that here that tests were sufficiently focused on a single characteristic that it was unlikely that the subjects’ criteria changed from test to test. The test results with adjusted means are shown in Fig. 14. With the exception of the case where $B = 25$ Hz, the major decrease in the tonality happens between 10 and 50 Hz cut-off frequencies, after which the tonalness is predominantly controlled by the modulation range ($B$). There is some evidence that the tonalness may increase at the highest cut-off frequency (200 Hz) for the modulation ranges of 50, 75 and 100 Hz, but the differences between the results at 100 Hz and 200 Hz are of the same order as the variations in results between tests. When the modulation range is small the tonalness remains high even though the frequency variation is fast.

C. Metric performance with the frequency modulated sounds

In this experiment the noise floor was provided by limitations in the instrumentation used in the subjective tests and was at -5 dB. The tone-width in the ANSI standard is defined as the intersection of a tonal feature and the noise floor. Thus, $\Delta f$ for most of the sounds is bigger than $\Delta f_{\text{tot}}$. This leads to problems when calculating $W_f$ in Eq. (1) and $\Delta L_f$ in Eq. (2). There were similar problems encountered when trying to calculate the Tonal Audibility ($\Delta L_t$) in the Joint Nordic Method. Aures’ model is made up of four terms, the first three account for bandwidth ($W_f$), center frequency ($f_c$)
and prominence ($w_p$) and the last term ($w_{Gr}$) is a function of the additional loudness provided by the tonal components and the total loudness of the signal. The Tone-to-Noise metric is most closely related to the $w_p$ and $w_{Gr}$ terms in the Aures' Tonalness metric which are shown in Figs. 15(a) and (b), respectively. The Aures' Tonalness metric calculations are shown in Fig. 15(c). The Aures' Tonality is not highly correlated with the subjective response data shown in Fig. 14 ($R^2 = 0.64$).

The poor agreement between the metrics and the subjective response data can be attributed to the calculation of the bandwidth from the spectra. It was hypothesized that the degree to which the frequency variation can be tracked should be factored into the bandwidth calculation. Taking shorter time-segments to produce spectral estimates would lead to spectral smoothing due to poor resolution, and lack of averaging would lead to high variability in the spectra, hence some other approach to estimating bandwidth is required. Therefore, the instantaneous frequency time histories were filtered to remove the components that could be tracked. This was achieved by passing these time histories through a 6 dB per octave roll-off high pass filter. The sound signals were then reconstructed with only the higher frequency non-trackable variation present and the spectra were regenerated. All 3 metrics were calculated from the spectra of the modified signals and the correlation between these and the subjective response values was examined. The coefficient of determination ($R^2$) was calculated for several high pass filter cut-off frequencies (5 Hz to 55 Hz); these are shown in Fig. 16. The best correlation occurred when the cut-off frequency was 25 Hz. The spectra generated from the modified signals for this case are shown in Fig. 17 and the Aures’ Tonalness values are shown in Fig. 18(a). The Tonal Audibility and Tone-to-Noise Ratio were also calculated from the modified spectra, these are shown in Figs. 18 (b) and (c), respectively.

When the modified spectra were used the correlation between Aures’ Tonalness and the subjective response data increased significantly ($R^2 = 0.88$). The $R^2$ value for the Tone Audibility was 0.82 and for the Tone-to-Noise Ratio was 0.82. Recall that before signal modification it was not possible to calculate these two metrics. Here the noise was defined to be all parts of the signal not associated with the “spiky” features centered at 700 Hz (see Fig. 17). This is slightly different to the approach described in the Joint Nordic Method for determining the noise contribution in the critical band, but is in the spirit of that calculation.
4. SUMMARY

Several metrics have been developed in order to model the perception of tonal components or their effect on people. Although each of these metrics are useful for a subset of tonal sounds, none of them work well on all types of tonal components. With the exception of Aures’ Tonalness metric, the other metrics measure the tonal prominence above the noise. It can be argued that these metrics are more signal based rather than perceptually based, though perceptual elements have been incorporated by how the metric is used. For example, in the Joint Nordic Method, the Tonal Audibility metric is used to apply penalties to average sound pressure measurements in a manner that reflects the added nuisance that is due to the presence of the tone in the noise beyond the increase in level that it contributes. Surprisingly, the Tonal Audibility does not have to be high to produce a large penalty, but the penalty saturates after the Tonal Audibility reaches 10 dB. However, the tonalness of these sounds, as predicted by Aures’ model, increases well beyond this level of Tonal Audibility as the tone power grows in relation to the noise power, but this additional tonalness may not impact annoyance. Further research is required to fully understand the relationship between tonalness and annoyance.

From the results of the subjective tests described in this paper, it can be concluded that roll-off rates of tonal features in spectra affect perception of the tonalness of sounds, and so a combination of bandwidth and roll-off rate must be used when estimating tonalness. This effect may be related to downwards frequency masking effects in the human auditory system. A modification of Aures’ metric is proposed, but further research is required to fine-tune this. It has also been shown that trackable nonstationary behavior of frequency modulated signals...
tones leads to difficulties when using tonal metrics that are derived from estimated spectra. The problem is primarily one of bandwidth estimation, although estimation of the noise contribution can also be problematic due to the wide region of the spectrum affected by the presence of the moving tone. A methodology to remove the trackable frequency component is proposed as a means to produce a more realistic measure of bandwidth. While results are promising, this too requires further research.

5. REFERENCES


