Comparison of listening test methods to evaluate tonality perception

Clement DENDIEVEL[1]; Antoine MINARD[1]; Christophe LAMBOURG[1]; Patrick BOUSSARD[1]
[1] GENESIS, France

ABSTRACT
Tonality is a perceptual attribute associated with tonal components (spectral emergences) present in many types of noise (aircraft, air conditioning systems, electric car ...). Several factors influence the perceived tonality (tone frequency, tone emergence level, etc). The frequency bandwidth of the background noise is one of those factors. Following an assumption of masking theory, most tonality indicators only take into account a critical bandwidth of noise around the emerging tone.
From a more practical standpoint, there are different listening test methods to study the impact of a given factor on perceived tonality (absolute evaluation, adaptive method ...). The duration of a listening test and the accuracy of perceptual results partly depend on the chosen method. In this study, 18 sounds, with varying bandwidth of background noise, were assessed by several subjects using three listening test methods: an absolute evaluation method, a two-alternative forced choice method and an iso-tonality equalization method. The iso-tonality equalization method provides the best compromise between the accuracy of the results and the listening test duration. Perceptual results also suggest that noise content outside the critical bandwidth around the tone should be considered when evaluating the tonality.

Keywords: Tonality, Perception, Listening test methods

I-INCE Classification of Subjects Number(s): 63.7, 69.3

1. INTRODUCTION

In many cases (aircraft noise (1), air conditioning (2), computer (3), train, electric car ...) tonality is perceived as inconvenient (4). The presence of tonal emergence(s) in noise can also be beneficial when designing car engine noise (5, 6). It is also an important functional factor in the perception of alarm sounds in background noise.

1.1 Factors influencing tonality perception

Several factors influence tonality perception. Some studies prove that perceived tonality increases with tone frequency and with tone emergence level (4, 7, 8, 9). Additional tones in noise influence tonality perception. The higher the number of tones, the higher the tonality is perceived.

Narrow bands of noise can also be perceived as tonal. It is experimentally proven that tonality perception decreases when increasing the spectral width of the tone (7, 8, 10, 11).

If the spectral width of tones is a factor that has been studied, the bandwidth of the background noise does not. In addition, some tonality indicators take into account in their calculations only a critical band of noise around the tone. These elements motivate the study of the potential influence of the noise bandwidth on the perceived tonality.

1.2 Listening test methods to study a factor's influence on tonality perception

Different methods of listening test are used to study the impact of a factor on the perceived tonality: paired comparison (9, 12), direct scaling (8), semantic differentials (13), or adaptive methods with a 2AFC response paradigm (2 Alternative Forced Choice). Two-AFC adaptive methods were originally used for threshold measurement: the threshold is determined after a number of trials between which the stimulus level is adjusted based on the previous responses of the subject. This type of method was

[1] cdendievel@optis-world.com, aminard@optis-world.com, clambourg@optis-world.com, pboussard@optis-world.com

[2] http://psychtoolbox.org/

then adapted to assess tonality perception (7, 14, 15).

2. EXPERIMENT

2.1 Stimuli

Eighteen stimuli are assessed, that correspond to all possible combinations of 3 tone frequencies (f_{tone}), 3 tone emergence levels (ΔL_{tone}) and 2 noise bandwidths (W_{noise}) (Table 1). Stimuli have a duration of 2 seconds. The tone emergence level is calculated as a Tone-to-Noise Ratio (16) – TNR –, which is the difference between the level of the tone and that of the critical band of noise centered on its frequency. TNR values of 5, 10 or 15 dB are considered. Noise bandwidths are expressed in critical band rate or Bark calculated according to Zwicker's model (17). Background noises are extracted from a Uniform Exciting Noise (UEN). A 6th-order Butterworth filter is used to create the desired spectral widths.

Table 1 – Experimental design

f_{tone} , Hz	ΔL_{tone} , dB	W_{noise} , Bark
600	5	2
1200	10	8
2400	15	

Some listening test methods use reference sounds (see 2.2.2 and 2.2.3). The reference sound is a UEN [20Hz - 20kHz], to which a tone is added ($f_{tone,ref}$ = 1 kHz). The loudness of 9 reference sounds (covering a large range of emergence levels) and of the 18 stimuli is experimentally equalized by 9 subjects.

2.2 Methods

Three listening test methods are implemented in order to study the influence of noise bandwidth on tonality perception.

Before starting the stimuli evaluation, each subject is given written instructions, describing the context of the study and the task to perform. He or she then receives a training explaining what tonality is. This training is followed by two basic and simple exercises allowing the experimenter to validate that each subject understood well this concept. Following this first phase, three tests are performed.

2.2.1 Test 1: direct scaling

During this test, each subject listens to the stimuli one by one in a random presentation order, and then indicates the level of perceived tonal intensity by positioning a cursor on a continuous scale (18) with 5 adjectives ("very low", "low", "medium", "strong", "very strong", translated from the French "très faible", "faible", "moyenne", "forte", "très forte" – see Figure 1).

For each sound, the result of the direct scaling method is a rating between the arbitrary values of 0 and 10 (associated to each end of the rating scale).

Figure 1 – Direct scaling test interface

2.2.2 Test 2: adaptive method (2AFC 1down / 1up)

During this adaptive test, subjects listen to two sounds and have to choose which one has the strongest tonal intensity (see Figure 2). One sound is one of the assessed stimuli, the other is a reference sound (see 2.1). At each trial, according to the previous response, the tone emergence level of the reference sound is either increased or decreased by a fixed step of 1 dB, in order to converge towards the tonal intensity of the assessed stimulus.

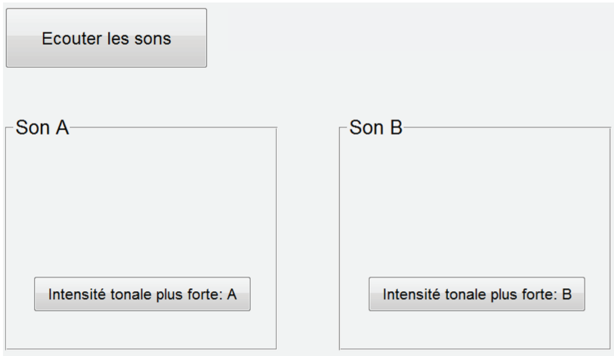
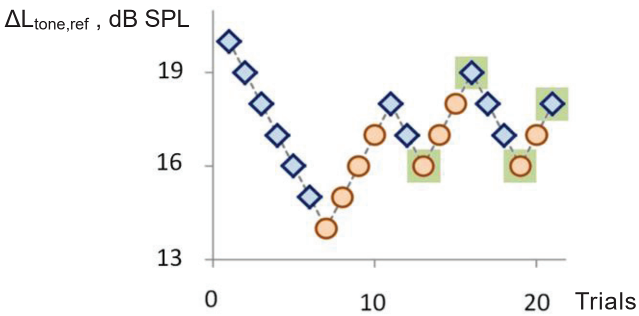


Figure 2 – 2AFC test interface

The test employs a 2AFC response paradigm and a 1down/1up adaptation rule. Each stimulus assessment ends after 6 inversions of the direction of adjustment (increase/decrease of $\Delta L_{tone,ref}$). Interleaved tracks are used and stimuli are presented in a random order. The tone emergence level of the reference sound ($\Delta L_{tone,ref}$, in dB) corresponding to the tonal intensity of the assessed stimulus is thus obtained for each sound. This emergence level is calculated by averaging the values obtained at the latest four inversions (see Figure 3).



◇ : the subject answers that the reference sound has the strongest tonal intensity,
○ : the subject answers that the tested stimulus has the strongest tonal intensity,
◇ or ○ : inversion (extremum) taken into account in the average of the reference emergence level.

Figure 3 – Obtained 2AFC result example

The loudness of the assessed stimuli and of the reference sounds is kept constant. To do so, a gain corresponding to linearly interpolated value from section 2.1 is applied at each trial to the reference sound.

2.2.3 Test 3 : Iso-tonality equalization

During this test, subjects adjust the tonal intensity of the reference sound (sound A) to match that of the assessed stimulus (sound B) (see Figure 4). The tonal intensity of the reference sound is adjusted with a slider that controls the emergence level of the tone ($\Delta L_{tone,ref}$, in dB). The tonality of the test stimulus (sound B) is assessed as the tone emergence level of the reference sound when the two have the same tonal intensity. It corresponds to the same output as the 2AFC method.

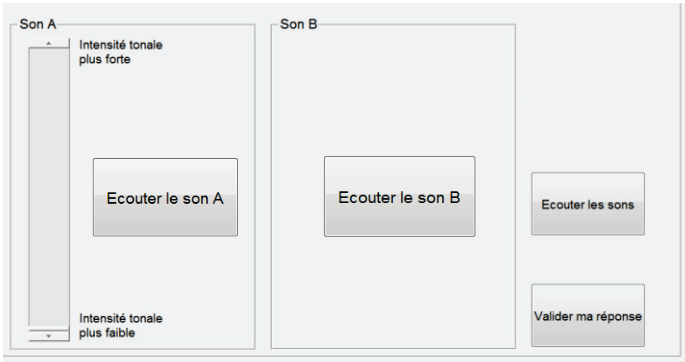


Figure 4 – Iso-tonality test interface

The loudness of the reference sound is not modified when the emergence level of the tone is adjusted, as for the 2AFC method (see 2.1 and 2.2.2).

2.3 Participants

Eighteen subjects (9M/9F, aged from 22 to 55) participated in the experiment for which they were paid. No subjects mentioned a major hearing problem. Ten participants were working in a field related to acoustics or audio.

2.4 Apparatus

The experiment took place in GENESIS' listening room. A Focusrite Scarlett 2i2 sound card and open headphones Sennheiser HD 650 were used. Three interfaces were specially programmed in Matlab R2011b. These interfaces support sounds' playback (Psychtoolbox library [2] and its PsychPortAudio function) and response input and storage for the three methods.

3. PERCEPTUAL RESULTS

A 3-way ANOVA with repeated measures was applied on the results of each method. In each case the 3 factors are f_{tone}, ΔL_{tone}, and W_{noise} (see Table 1). Resulting significant p values are displayed in Table 2 (in bold). These results are detailed and illustrated in the following subsections.

Table 2 – Perceptive results, ANOVA (pVal)

Parameter	Direct Scaling	Equalization	2AFC
ΔL_{tone}	**0.000**	**0.000**	**0.000**
f_{tone}	**0.000**	**0.000**	**0.002**
$f_{tone} * \Delta L_{tone}$	0.507	0.284	**0.003**
W_{noise}	**0.000**	**0.000**	**0.000**
$\Delta L_{tone} * W_{noise}$	0.155	**0.000**	**0.000**
$f_{tone} * W_{noise}$	0.436	**0.006**	**0.025**

[2] http://psychtoolbox.org/

3.1 Tone emergence level (ΔL_{tone}) influence on tonality perception

Perceived tonality increases when tone emergence level increases (Figure 5). This influence is significant for any method used (Table 2). On average, when ΔL_{tone} increases by 5 dB, ratings increase by 1.2 units for the direct scaling method, and $\Delta L_{tone\ ref}$ increases by around 8 dB for both the 2AFC and the iso-tonality equalization methods.

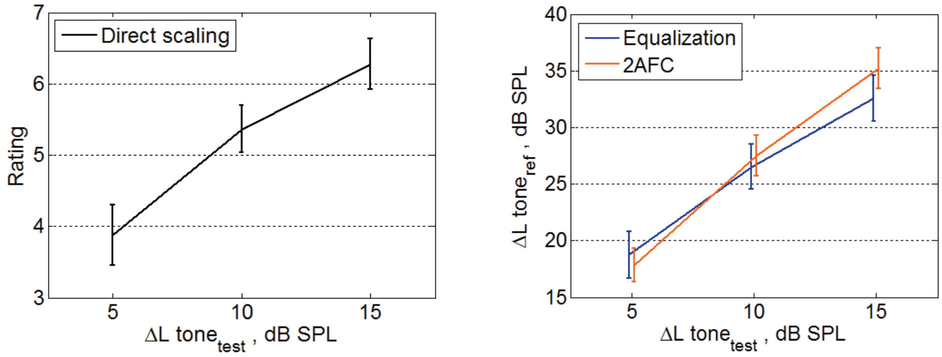


Figure 5 – Tone emergence level influence on tonality perception

3.2 Tone frequency (f_{tone}) influence on tonality perception

Perceived tonality increases when tone frequency increases (Figure 6). This influence is significant for any method used (Table 2). On average, ratings increase by 0.7 units for the direct scaling method, and $\Delta L_{tone\ ref}$ increases by 3 dB for both the 2AFC and the iso-tonality equalization methods.

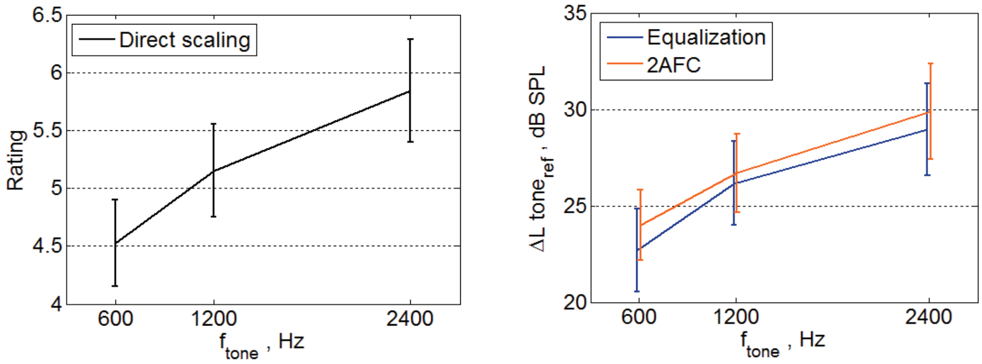


Figure 6 – Tone frequency influence on tonality perception

3.3 Background noise width (W_{noise}) influence on tonality perception

Perceived tonality decreases when noise bandwidth increases (Figure 7). This influence is significant for any method used (Table 2). On average, when the noise bandwidth go from 2 to 8 Bark, ratings decrease by 1.3 units for the direct scaling methods, and $\Delta L_{tone,\ ref}$ decreases by 9 dB for both the 2AFC and the iso-tonality equalization methods. These values show that the influence of noise bandwidth is as strong as that of tone emergence level for the studied sounds.

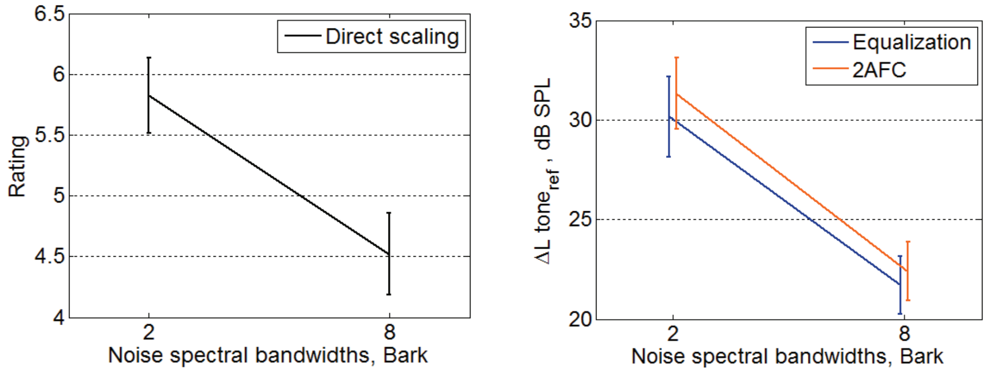


Figure 7 – Noise spectral width influence on tonality perception

3.4 Parameter interactions

The ANOVA results show significant interactions between some parameters for the equalization and the 2AFC methods, whereas no interaction is shown for the direct scaling method. These interactions are not well identified when analyzing the results into more details.

4. COMPARISON OF THE METHODS

4.1 Correlations of the results

Averaged perceived tonality values are calculated for each stimulus and for each method. Pearson product-moment correlation coefficients are then calculated. The highest correlation coefficient is found between the equalization and the 2AFC results (R=0.961, 16 degrees of freedom, see Figure 8 and Table 3). This is probably due to similar nature of the scale obtained with these 2 methods. Overall all correlation coefficients are very high. This means that all 3 methods give roughly the same mean results.

Table 3 – Correlation coefficient between the results of the 3 methods

R (16 degrees of freedom)	Direct scaling	Equalization	2AFC
Direct scaling	-	0.918	0.942
Equalization	0.918	-	0.961
2AFC	0.942	0.961	-

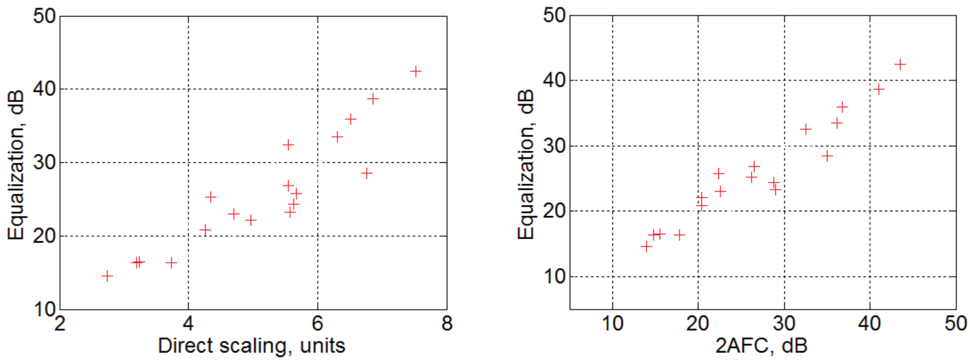


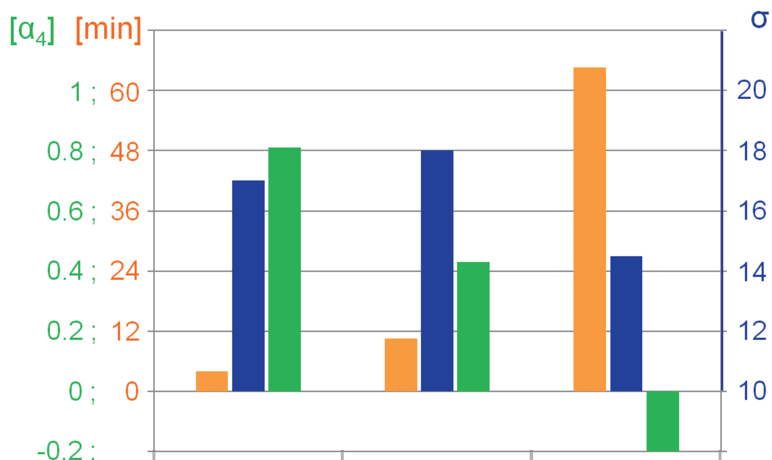
Figure 8 – Scatter plot of the mean results of the methods:
Equalization vs Direct scaling (left plot), Equalization vs 2AFC (right plot)

4.2 Methods' durations, standard deviations and kurtosis

Given that the direct scaling results are not on the same type of scale as the two other methods, all data need to be converted to a common scale to be compared. The results of each method are converted to a 0-100 scale. The following simple formula is used (the maximum and minimum values for a method are chosen between 18 subjects * 18 stimuli = 324 values):

$$result_{0-100}(i) = \frac{result(i) - min_{method}}{max_{method} - min_{method}} * 100$$

Then, the standard deviation (STD) is calculated for each stimulus and for each method, and is further averaged over the 18 stimuli. The resulting average STD values for the 3 methods, as well as the average kurtosis values obtained in the same manner, are displayed in Figure 9. This figure also shows the mean duration of each method.



	Direct scaling	Equalization	2AFC
Duration	4min	10min	65min
STD	17.0	18.0	14.5
Kurtosis	0.81	0.43	-0.20

Figure 9 – Methods' duration and accuracy comparison

The 2AFC method has the longest duration, but seems to be the most accurate method. On the opposite, the direct scaling method has the shortest duration and the less accurate method (high STD and kurtosis). The difference between the mean STD values of the direct scaling and equalization methods is probably not significant, but the 2AFC method appears to have a smaller mean STD value. On a more general standpoint, the equalization and 2AFC methods have the advantage of outputting results on a physical scale (dB) which makes easier their interpretation on an absolute level.

For these reasons, the equalization method appears to offer a good comprise between duration (6.5 times faster than the 2AFC method), accuracy and ease of interpretation.

4.3 Participants' feeling

After each method, participants are asked to express their feeling on the task to perform.

4.3.1 Direct scaling

Most of the participants found this method quite hard, even though its duration is short. The major difficulty for them was to give an answer without being able to compare the stimuli together. There appears to be 3 rating profiles:

- ranking around the medium graduation,
- rating over only a part of the scale, either the lower or the higher part,
- rating over the full scale.

These different profiles could explain that the direct scaling method has the highest mean kurtosis, as results have a "heavy-tailed" distribution (i.e. values are bounded by the ends of the rating scale).

4.3.2 Equalization

Most of the participants said that the equalization method is easy to handle, and that assessing stimuli is much easier when they can compare the stimuli. The major difficulty for them was to equalize two sounds that do not have the same noise (20-20kHz noise bandwidth for the reference sound, 2 or 8 critical bandwidths of noise for the assessed stimuli). It was the participants' preferred method.

4.3.3 Two-AFC

Every participant said that the 2AFC method is very tiresome and difficult. Firstly, the method is very long (more than 1 hour long on average). Secondly, it was really hard for some participants to select the most tonal sound when the 2 presented sounds had the same perceived tonality. Participants may have gotten frustrated because of that.

5. TONALITY INDICATORS' COMPARISON

One aim of this study is to address whether tonality indicators take into account the influence of the noise bandwidth. The considered indicators are:

- Maximum Prominence Ratio (PR) and Tone-to-Noise Ratio (TNR) values (16)
- "Mean Difference" of the DIN45681 standard (19)
- "Tonal Audibility" of the annex C of the ISO1996-2 standard (20)
- Aures' model of tonality, averaged over time (21).

Nine stimuli were compared: 3 frequencies (600Hz, 1200Hz and 2400Hz) and 3 noise bandwidths (2, 4, and 8 Bark). Tone emergence levels are fixed at 10 dB. These stimuli have duration of 60 seconds (required minimum duration in the ISO1996-2 standard). Results are displayed on Figure 10 for f_{tone} = 1200Hz (results' shapes are similar for f_{tone} = 600 and 2400 Hz).

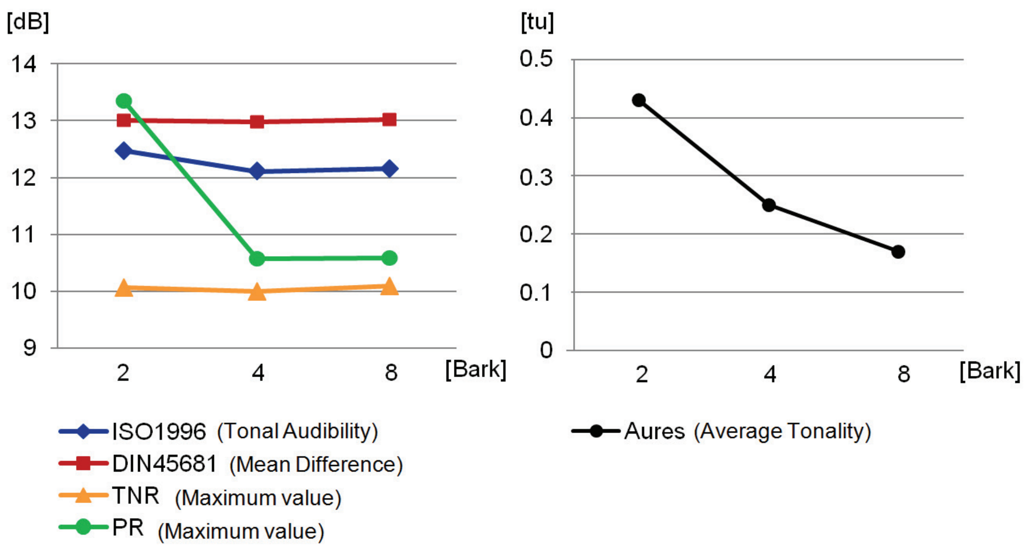


Figure 10 – Tonality indicators comparison, f_{tone} = 1200 Hz, ΔL_{tone} = 10 dB

According to the listening test results, perceived tonality decreases when background noise bandwidth increases (Figure 7). ISO1996, DIN45681, and TNR indicators do not take into account this particular influence. As the PR indicator calculates the energy ratio between the critical band of the tone and the two adjacent critical bands, it is logical that the PR value is higher when the noise has a bandwidth of only 2 critical bands. Over a noise bandwidth of 3 Bark, PR values remain constant (all else being equal).

Aures' tonality is the only indicator that takes into account the influence of the noise bandwidth, by applying a weighting based on the relative loudness of the tone and the noise, thus accounting for the whole spectrum of the noise.

6. SUMMARY

In this study, 3 listening test methods (direct scaling, 2AFC 1down/1up and iso-tonality equalization) were implemented in order to study the influence of the background noise bandwidth on tonality perception. Eighteen stimuli were assessed by 18 participants for each method.

The results concerning the influence of the tone emergence level and the tone frequency are in line with the literature: perceived tonality increases with both tone emergence level and frequency.

It appeared that the noise bandwidth has an influence on tonality perception: perceived tonality decreases when the noise bandwidth increases.

The 3 listening test methods gave similar results. The iso-tonality equalization method seems to offer the best compromise between duration, accuracy, ease of interpretation and participant feeling.

As far as tonality indicators are concerned, Aures model is the only one that takes into account the influence of the noise bandwidth.

This study proposes that tonality should be estimated by an indicator based on the relative loudness of the tonal content and the noise. Some non-standardized indicators have already been tested in this way in the literature. Further listening tests and methods of tonality calculation will be carried out to improve tonality estimation.

REFERENCES

1. Brocolini L., Parizet E. Influence of tonal components on the unpleasantness of airplane noise at take-off, Proceedings of Internoise 2015, San Francisco, USA, 2015.
2. Yamaguchi T., Minorikawa G. Kihara M., Study on evaluation method of the pure tone for small fan, The 22nd International Congress on Sound and Vibration, Florence, Italy, 2014
3. Sottek R. Progress in calculating tonality of technical sounds, Internoise 2014, Melbourne, Australia, 2014
4. Parizet E., Bolmont A., Fingerhuth S. Subjective Evaluation of tonalness and relation between tonalness and unpleasantness, Internoise 2009, Ottawa, Canada, 2009
5. Boussard P., Molla S., Orange F. Comprehensive Process for Car Engine Sound Design: from signal processing to an audio system integrated in the vehicle, Internoise 2012, New-York City, USA, 2012
6. Boussard P., Khurana R., Guyader G., Orange F., Orzechowski J. Implementing digital engine sound enhancement techniques to define and refine vehicle interior sound image/quality, Internoise 2013, Innsbruck, Austria, 2013
7. Vormann M., Verhey JL, Mellert V., Schick A. Subjective rating of tonal components in noise with an adaptative procedure, 2002
8. Sottek R. A new hearing model approach to tonality, Internoise 2013, Innsbruck, Autria, 2013
9. Vormann M., Meis M., Mellert V., Schick A. A new approach for the evaluation of tonal noise (tonality), Psychophysics, Physiology and Models of Hearing (pp. 109-112), 1998
10. Taghipour A., Edler B., Armirpour M., Herre J. Dependency of tonality perception on frequency, bandwidth and duration, 21st International Congress on Acoutics, Montreal, Canada, 2013
11. Hastings A., Davies P. The Effect of Attenuation Rate and Peak Bandwidth of Tonal Components on Perceived Tonalness, First International Speech and Communication Association ITRW, Mont-Cenis, Germany, 2003
12. Hansen H., Weber R. The influence of tone length and S/N-ratio on the perception of tonal content: an application of probabilistic choice models in car acoustics, The Acoustical Society of Japan, 2008
13. Hansen H., Weber R., Letens U., Quantifying tonal phenomena in interior car sounds, Forum Acusticum, Budapest, Hungary, 2005
14. Oetjen A., Volk P., Van de Par S. Perception of stationary and transient signals, Euronoise 2015, Maastricht, Netherlands, 2015
15. Verhey JL, Heise SJ Suprathreshold perception of tonal components in noise under conditions of masking release, Acta Acustica united with Acustica, 2012
16. ECMA-74, Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment, 13th edition, 2015
17. Fastl H., Zwicker E. Psychoacoustics: Facts and Models. Berlin, Germany: Springer, 3rd edition, 2007
18. Parizet E., Hamzaoui N., Sabatie G. Comparison of some listening test methods : a case study, Acta Acustica united with Acustica, 2005
19. DIN 45681:2005-03, Acoustics, Detection of tonal components of noise and determination of tone adjustment for the assessment of noise immisions, 2005
20. ISO1996-1 Acoustics, Descriptions, measurement and assessment of environmental noise , Part 2 : Determination of environmental noise levels, 2007
21. Aures W., A model for calculating the sensory euphony of various sounds, *Acustica*, **59**, pp. 130-141, 1985