forum acusticum 2026
8-12 September · Graz, Austria

MAGNITUDE OF TONAL CONTENT AS FUNCTION OF LEVEL

Andre Fiebig[1], Felix Hochbaum[1], Florian Doleschal[2], Jesko Verhey[2], Julia Treichel[3]

[1] *Department of Engineering Acoustics, TU Berlin, Einsteinufer 25, 10587 Berlin, Germany*
[2] *Department of Experimental Audiology, Otto von Guericke University Magdeburg, Leipziger Str. 44, 39120 Magdeburg, Germany*
[3] *Umweltbundesamt (German Environment Agency), Wörlitzer Platz 1, 06844 Dessau-Roßlau, Germany*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

There is an increasing number of sound sources in our environment emitting sounds containing tonal components. Their perception is commonly assessed by measuring the magnitude of tonal content, also known as the tonality of the sound. While several studies investigated the impact of the level of tonal content relative to the level of the noise background, little is known about the effect of the overall level on the magnitude of tonal content. This study presents data collected in three German laboratories (Berlin, Magdeburg, Dessau) on the effect of a change in overall level of the sound (keeping the tone-to-noise ratio constant) on the magnitude of tonal content. Technical and artificial sounds were used. The latter sounds were partly basic stimuli (pure tone in noise) and partly stimuli, which were motivated by real sounds of electric vehicles. In total, 77 listeners participated in the experiment; none of them reported hearing impairment. For most stimuli, the measured magnitude of tonal content increases with level. This is in contrast to tonal methodologies such as tone-to-noise ratio and prominence ratio that were defined in national (DIN 45681) and international standards (ISO/TS 20065).

Keywords: *psychoacoustics, tonality, modeling*

1. Introduction

In our modern environment, we are often exposed to sounds, which contain tonal components. Examples are sounds from wind turbines, trains or heat pumps. Sounds in the vehicle interior also often contain tonal components, originating from, e.g., the gearbox, the electrically operated window regulators or electrically folding door mirrors. Sounds where the tonal components are audible are commonly rated as being more annoying. Several standards attempt to quantify the perception of tonal components in noise (e.g. IEC 61400-11:2006 [1]; ANSI S1.13, 2005 [2]; ISO/TS 20065:2022 [3], ECMA 418-1:2024 [4], ECMA 418-2:2025 [5]).

In Germany, DIN 45681:2005 [6] may be used to characterise not only the perception of tonal components, but also its higher annoyance than sounds without audible components, where penalties up to 6 dB are applied in case of detected strong tonal components. It was observed that for pure multi-tone sounds the penalty levels were found to be even above 6 dB indicating that for these stimuli the DIN 45681 standard potentially underestimates the required penalty level [7].

As in several other national and international standards mentioned above, sounds are analysed spectrally, prominent tonal components are detected, and their level above the masked threshold is determined. The component with the highest level above the masked threshold is then used to determine a tone adjustment to account for the higher annoyance of sounds containing audible tonal components. In general, standards assume that the perception of the audible components depends solely on the level above the masked threshold, whereas the overall level of the sound is not considered. In contrast, a dependence on the overall sound pressure level is rarely regarded as relevant; see as an exception ECMA418-2.[5]

The present study investigates if this latter assumption is valid by measuring the magnitude of the tonal content of sounds as a function of their level. The magnitude of tonal content or tonality is the quantity that is associated with the perception of the tones in noise (cf. [8], for a review) Artificial sounds as well as real recordings were used. The magnitude of tonal content is measured using absolute magnitude estimation procedure with categories. Such an approach was already used, e.g., in [9] to assess the magnitude of the tonal content of artificial cabin interior sounds of electric vehicles. In the present study, a subset of

Corresponding author: *andre.fiebig@tu-berlin.de.*

their sounds was also used.

2. Method

To investigate the effect of the overall sound pressure level on the perceived magnitude of the tonal content, a listening experiment was designed and conducted as a round-robin study across three sites. Experimental data were collected at Umweltbundesamt (German Environment Agency) in Dessau-Roßlau, Otto-von-Guericke-Universität Magdeburg and Technische Universität Berlin. Participants rated a selection of signals presented at different total levels (see table 1) for their magnitude of tonal content on a unipolar 9-point rating scale ranging from “not tonal” to “extremely tonal”.

2.1. Stimuli

Table 1: Stimulus sets and the corresponding listening test playback levels.

Stimulus Set	Playback Level	
	L_p	L_A
12-Fensterheber	40 dB [SPL]	36 dB [A]
	60 dB [SPL]	56 dB [A]
	80 dB [SPL]	76 dB [A]
5-CVT-Getriebe	69 dB [SPL]	55 dB [A]
	89 dB [SPL]	75 dB [A]
cd_1_25_0.2_10	79 dB [SPL]	50 dB [A]
	99 dB [SPL]	70 dB [A]
cd_1_25_0.4_10	79 dB [SPL]	50 dB [A]
	99 dB [SPL]	70 dB [A]
cd_1_25_unmod	79 dB [SPL]	50 dB [A]
	99 dB [SPL]	70 dB [A]
cd_2_25_unmod	79 dB [SPL]	50 dB [A]
	99 dB [SPL]	70 dB [A]
cd_updown_25_unmod	79 dB [SPL]	50 dB [A]
	99 dB [SPL]	70 dB [A]
zz noise only	40 dB [SPL]	36 dB [A]
	60 dB [SPL]	56 dB [A]
	80 dB [SPL]	76 dB [A]
zz sound 1000_42	40 dB [SPL]	36 dB [A]
	60 dB [SPL]	56 dB [A]
	80 dB [SPL]	76 dB [A]
zz sound 1000_54	41 dB [SPL]	38 dB [A]
	61 dB [SPL]	58 dB [A]
	81 dB [SPL]	78 dB [A]
zz sound 500_48	40 dB [SPL]	36 dB [A]
	60 dB [SPL]	56 dB [A]
	80 dB [SPL]	76 dB [A]

A selection of 11 signals was made to serve as pool of stimuli, including synthetically created as well as recorded technical sounds. Technical sounds with tonal content were a window lift (ger: Fensterheber) and CVT transmission sound (ger: CVT-Getriebe). Both sounds were originally used in a study on the magnitude of tonal content presented by Letens at a workshop of the technical committee on vehicle acoustics of the German Acoustical society (DEGA) in 2008 [10]. These stimuli were also used in [11].
Five of the selected signals were subsequently generated at three distinct levels separated by 20 dB (e.g. L_p of 40 dB, 60 dB and 80 dB), whereas the remaining six were adjusted to two level settings with the same 20 dB spacing, resulting in a total of 27 different stimuli.
The stimuli with labels starting with “zz” consist of a broadband pink noise which is either presented alone (“zz noise only”) or together with a sinusoid. The signal to noise ratio is encoded in the last number of the signal name. The second-last number indicates its frequency in Hertz. The signals with labels starting with “cd” are simulated coast-down signals. Except for the signal labelled up-down, they were already used in [9]. The labelling is described in detail in that publication. Briefly, the coast-down signal consisted of a background noise and either one or two sweeping components (indicated by the first number in the label). This were either unmodulated (labels containing “unmod”) or modulated at a rate in Hertz given by second-last number (i.e., 0.2 or 0.4 Hz). For the signal containing “updown” tone frequency first decreased and then increased over time. For all other coast-down signal, only sweeps with decreasing frequency over time were used. The stimuli were 2 s (technical sound, sounds starting with “zz”) or 5 s (signals starting with “cd”) in duration. Table 1 lists all stimulus sets and the corresponding playback levels in the listening experiment.

2.2. Apparatus

The experimental apparatus varied slightly between sites due to local circumstances and available equipment, but conditions were kept as similar as possible across all three locations. The same custom built MATLAB Graphical User Interface was used in all sites. Sennheiser HD 600 (Berlin) and HD 650 headphones (Magdeburg, Dessau) were paired with high-quality audio interfaces to ensure optimal audio playback. The playback levels were calibrated using Brüel & Kjaer type 4153 Artificial Ears to ensure identical playback levels across all locations. While audiometry cabins were available in Magdeburg and Berlin, a semianechoic chamber of quality class 1 (20 dB(A)) was chosen in Dessau-Roßlau.

2.3. Procedure

Participants were first informed by the experimenter about the purpose, procedure, and duration of the experiment, as well as the scientific use of the data, and asked for known hearing impairments. Thus, all participants were provided information about the study and consent was obtained from all participants; they agreed to the scientific use and publication of the results.
Then they were presented with a demonstration sound to familiarize themselves with the concept of magnitude of tonal content. The demonstration sound consisted of a tone–noise complex in which a 1 kHz sine tone smoothly faded in and out over a pink-noise base

during the 10 s signal interval. Participants could replay the sound as often as desired. When familiarized, participants could start the first of three experimental blocks by closing the introduction window. In each experimental block, participants were presented with all 27 stimuli in randomized order once. The signals could not be replayed, and an answer could not be changed when given. This ensured that the participants reported their first impression of the stimulus. In addition, the duration of the experiment was about the same for all participants. Breaks could be taken freely between blocks. With 27 ratings per block, a total of 81 ratings had to be made in the experiment as each stimulus was rated 3 times by each participant. The total duration of the experiment was summed up to 30 minutes. The experimental data collected during the experiment were anonymized. The experiments at the sites were carried out in the spring and summer of 2025.

2.4. Results

In total, 77 participants took part in the listening experiment across the three locations: n=32 in Dessau-Roßlau (DE), n=23 in Berlin (BE), and n=22 in Magdeburg (MD). All participants self-reported normal hearing. Each participant provided three ratings for each stimulus. In a first step, the repeated assessments were averaged to obtain a single, more stable estimate for each participant and signal in order to reduce random variability. Because the signal "zz noise only", that does not contain any tonal components but just stochastic noise, is conceptually non-tonal, an exclusion criterion for participants was defined. Participants who rated the "zz noise only" signal above 6 ("very tonal") on average were excluded from further analyses, resulting in a final sample of 68 participants (n=26 in Dessau-Roßlau, n=21 in Berlin, and n=21 in Magdeburg). For the remaining participants, the intra-rater-reliability was considered by calculating the absolute mean difference of repeated ratings per participant and then averaged over the entire sample per laboratory. It was observed that the mean intra-rater-reliability per laboratory was comparable: in DE, the absolute mean difference of repeated ratings was 0.84, in MD 0.83 and in BE 0.74 categories. The lowest observed intra-rater-reliability in terms of absolute mean difference of repeated ratings of all 68 participants was lower than 1.5 categories, which is why no further individuals were excluded.

To decide whether the remaining data of the three sites are eligible for pooling, a repeated-measures ANOVA was conducted with signal as the within-subject factor and site as the between-subject factor. Due to violated sphericity assumptions, Greenhouse-Geisser corrected values are reported for the within-subjects effect. While there was no between-subject effect on site ($F(2, 65) = 2.30, p = .109$), there was a highly significant within-subject interaction between signal and site ($F(15.62, 507.79) = 4.65, p < .001$). Given these differences between sites, pooling of the data was considered inappropriate in a statistical sense. The ratings of the magnitude of tonal content at the three sites for each stimulus set and the respective playback levels are presented as boxplots in Figure 1.

Although the data was not pooled across the laboratories, the correlations between the mean magnitude of tonal content ratings per signal collected at the different sites were found to be statistically significant on the basis of the Pearson correlation coefficient. The lowest correlation coefficient, 0.91, was observed between the BE and DE data. The other correlations between the average ratings of the magnitude of tonal content collected at the different sites were even higher and reached 0.95 between MD and BE. This means that the general rating behaviour regarding the magnitude of tonal content per signal was similar between the laboratories. Moreover, the general trend that a level increase leads to an increase of judged magnitude of tonal content is comparable between the locations with a few exceptions. In BE for a few signals, the level dependencies are less certain compared to MD and DE. To quantify differences in ratings between levels at each site and test signal more in detail, *Cohens D* [12] was calculated as a descriptive effect size measure. Inferential hypothesis tests (e.g., t-tests, ANOVAs) were omitted to avoid alpha inflation from the multiple comparisons that would be necessary and to emphasize practical relevance over statistical significance. The effect sizes for each level comparison at each site are presented alongside the mean rating differences in Table 2.

3. Discussion

Although the results indicate an overall level dependence of perceived magnitude of tonal content for most of the signals, the results are far from being conclusive. In particular, the small level effect for broadband noise only (signal "zz noise only") regarding the judged magnitude of tonal content, might suggest that the collected magnitude of tonal content assessments are confounded by annoyance. However, it may be argued that observed small effect sizes at the floor (or ceiling) of rating scales have to be interpreted with caution.

Thus, although for several signals large effect sizes were observed for the level dependence of the magnitude of tonal content, one may argue that the effect of total level on the magnitude of tonal content is slightly overestimated (part of it may just reflect an increase in annoyance as the level increases). In particular, the experimental results from Berlin showed for some signals less to no total level effect on perceived magnitude of tonal content. However, the debate over the extent to which one can distinguish between sensory dimensions -such as loudness or magnitude of tonal content- and annoyance has been ongoing for decades, with both sides having their supporters (see exemplarily [13]). Moreover, due to the convenience sampling method (recruitment of participants) used at the different sites sample related differences cannot be ruled out potentially relevant for the small differences

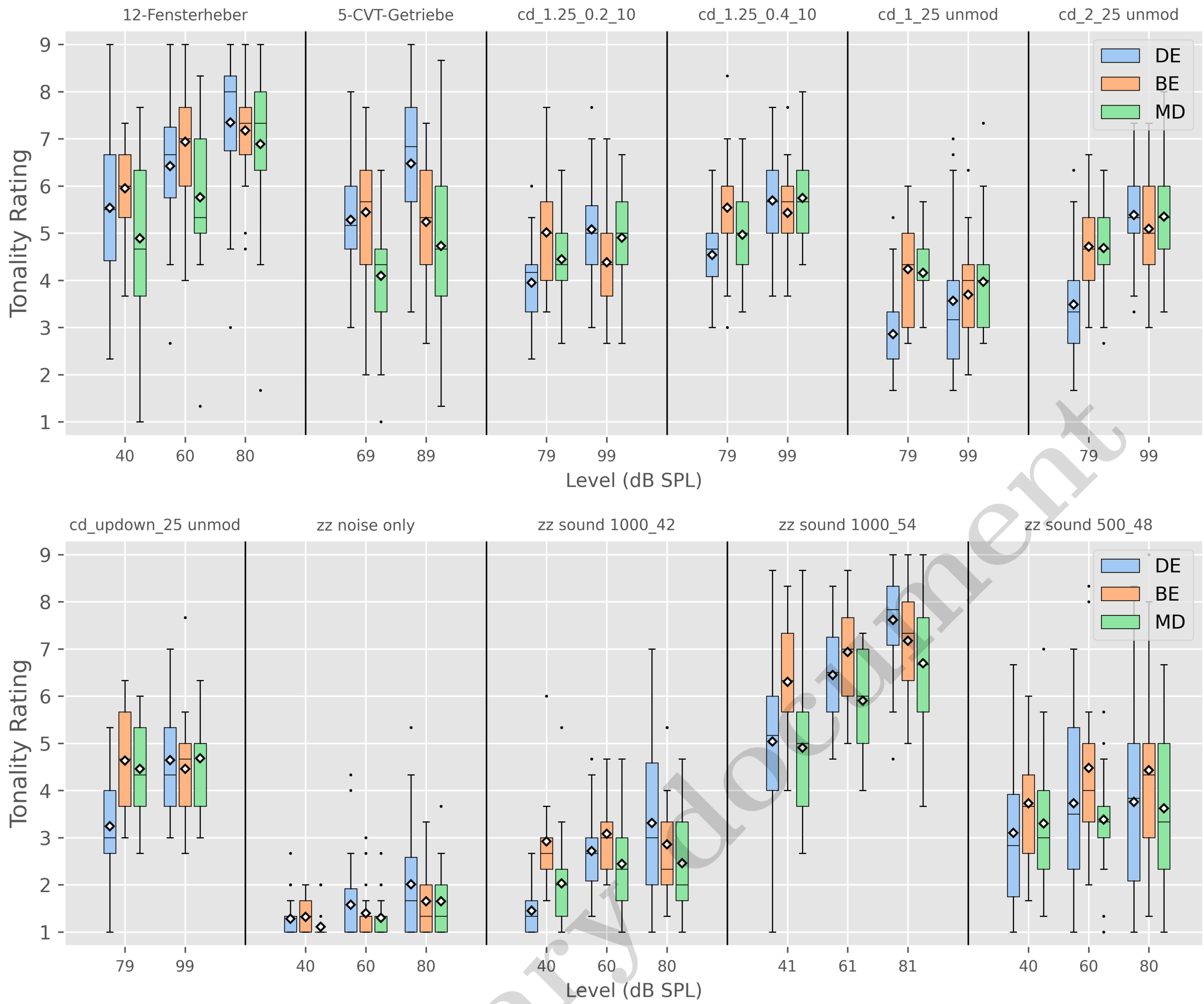

Figure 1: Boxplots showing the distribution of the magnitude of tonal content ratings for every site, stimulus set and playback level. Besides the boxes containing 50% of the data and the whiskers, the median (horizontal line) and arithmetic means values (diamond) are shown.

between the laboratories. For future experiments, it may be interesting to collect more data about sample characteristics to explore differences in tonality judgments in greater detail.

For example, Oetjen and van de Par did not observe a clear effect of total level on magnitude of tonal content judgments in an experiment using an adaptive staircase procedure to determine the point of subjective equality [14].

As established standards for the determination of magnitude of tonal content, such as tone-to-noise (DIN 45681) or prominence ratio (ECMA 418-1) are mainly based on the level of the tonal part (e.g. by determining a discrete tone level) compared to the level of the noise background (e.g by deriving the total critical band sound pressure level) and do not take into account the total level of the signal, the interpretation of the experimental results is of crucial importance and must be carried out carefully. Some experimental studies suggest that perceived magnitude of tonal components is strongly related to the partial loudness rather than the level of the tonal content of a sound ([8], [15] and [16]).

In contrast to standards that depend on a comparison of levels, methods and approaches more focused on the partial loudness of the tonal content might result in an increase in the magnitude of the tonal content with increase of the total level. For example, this model behavior is found in the ECMA 418-2 focussing on the estimation of tonal loudness, where an increase of the total level results in an increased magnitude of tonal content. Thus, further research is needed to determine whether the perceived magnitude of tonal components depends generally on the overall signal level, or whether, while an overall level effect on magnitude of tonal content may occur with certain signals, it does not hold in general.

Table 2: Paired Cohen's d_z and pairwise difference of mean values (Δmean) for all signal-level comparisons across sites. Descriptive effect size interpretation: small ≥ 0.2, medium ≥ 0.5, large ≥ 0.8.

signal group	comparison	DE		BE		MD	
		d_z	Δmean	d_z	Δmean	d_z	Δmean
12-Fensterheber	40 vs. 60 dB	0.640	0.885	0.922	0.984	0.640	0.873
	60 vs. 80 dB	0.982	0.923	0.313	0.238	0.718	1.127
5-CVT-Getriebe	69 vs. 89 dB	0.806	1.192	-0.196	-0.206	0.459	0.635
cd_1.25_0.2$_1$0	79 vs. 99 dB	1.074	1.128	-0.606	-0.635	0.427	0.460
cd_1.25_0.4_10	79 vs. 99 dB	1.155	1.154	-0.144	-0.111	0.974	0.778
cd_1_25 unmod	79 vs. 99 dB	0.449	0.705	-0.409	-0.540	-0.178	-0.190
cd_2_25 unmod	79 vs. 99 dB	1.753	1.897	0.349	0.381	0.645	0.667
cd_updown_25 unmod	79 vs. 99 dB	0.936	1.397	-0.129	-0.175	0.191	0.222
zz noise only	40 vs. 60 dB	0.460	0.295	0.140	0.079	0.657	0.190
	60 vs. 80 dB	0.386	0.436	0.420	0.254	0.656	0.349
zz sound 1000_42	40 vs. 60 dB	1.984	1.269	0.193	0.159	0.560	0.413
	60 vs. 80 dB	0.405	0.590	-0.294	-0.222	0.014	0.016
zz sound 1000_54	41 vs. 61 dB	0.853	1.410	1.005	0.635	0.852	1.000
	61 vs. 81 dB	1.045	1.167	0.252	0.238	0.521	0.794
zz sound 500$_8$	40 vs. 60 dB	0.500	0.628	0.712	0.746	0.055	0.079
	60 vs. 80 dB	0.015	0.026	-0.041	-0.048	0.171	0.238

4. Summary

An experiment was designed and conducted in three different laboratories to determine the effect of a change in overall sound pressure level on the magnitude of the tonal component. In general, the experimental results were similar between the laboratories in Dessau, Magdeburg and Berlin, even though, from a statistical standpoint, combining the results was not valid. For most of the signals the judged magnitude of tonal content increased with increasing the sound pressure level by 20 or 40 dB. In addition, an increase of the assessed magnitude of tonal content for a reference signal without prominent tonal components (broadband stochastic noise only (zz noise only)) was observed, although with very low values, a partial confounding of the magnitude of tonal content ratings by annoyance cannot be ruled out. Thus, further experiments have to determine the magnitude of the effect of the total level on perceived magnitude of the tonal component(s) in more detail.

5. Acknowledgments

The authors would like to thank all members of the DIN working group NA 001-01-10 AA 'Measurement techniques for psychoacoustics' for their contributions to the planning of experiments and the selection of signals.

References

[1] IEC, "Acoustic noise measurement techniques," International Electrotechnical Commission, Standard IEC 61400-11, 2006.

[2] ANSI/ASA, "Measurement of Sound Pressure Levels in Air," American National Standards Insitute, Standard ANSI 1.13, Jun. 2020.

[3] ISO, "Acoustics. Objective method for assessing the audibility of tones in noise — Engineering method," International Organization for Standardization, Standard ISO/TS 20065, 2022.

[4] ECMA international, "Psychoacoustic metrics for ITT equipment — Part 1 (prominent discrete tones)," ECMA international, Standard 418-1, Jun. 2024.

[5] ECMA international, "Acoustics - Determination of tonal components of noise and determination of a tone adjustment for the assessment of noise immissions," ECMA international, Norm 45681, Mar. 2005.

[6] DIN Deutsches Institut für Normung e.V., "Acoustics - Determination of tonal components of noise and determination of a tone adjustment for the assessment of noise immissions," Beuth Verlag, Norm 45681, Mar. 2005.

[7] S. Toepken, S. van den Par, and R. Weber, "Subjective evaluation of loudness and preference for noise containing audible tones," *J. Acoust. Soc. Am.*, vol. 137, p. 2320, Apr. 2015. DOI: https://doi.org/10.1121/1.4920466.

[8] H. Hansen, J. Verhey, and R. Weber, "The magnitude of tonal content. a review," *Acta Acustica united with Acustica*, vol. 97, pp. 355–363, May 2011. DOI: 10.3813/AAA.918416.

[9] F. Doleschal and J. L. Verhey, "Pleasantness and magnitude of tonal content of electric vehicle interior sounds containing subharmonics," *Applied Acoustics*, vol. 185, p. 108 442, 2022, ISSN: 0003-682X. DOI: https://doi.org/10.1016/j.apacoust.2021.108442. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0003682X21005363.

[10] U. Letens, "Ringversuch zur Tonhaltigkeit des DEGA- Fachausschusses für Fahrzeugakustik (FAFA), Workshop des FAFA in Herzogenrath," DEGA, Report, 2008.

[11] J. L. Verhey, A. Oetjen, and R. Weber, "Zeitlich kontinuierliche Bewertung der Tonhaltigkeit instationärer Geräusche," in *Proc. of DAGA*, Darmstadt, Germany, 2012, pp. 47–48.

[12] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," en, *Frontiers in Psychology*, vol. 4, 2013, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00863. Accessed: Apr. 25, 2024. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/abstract.

[13] W. Reichardt, "Subjective and objective measurement of the loudness level of single and repeated impulses," *J. Acoust. Soc. Am.*, vol. 47, pp. 1557–1562, Apr. 1970. DOI: https://doi.org/10.1121/1.1912087.

[14] A. Oetjen and S. van der Par, "Spectral prominence influencing the perceived strength of psychoacoustic measures," in *Proc. of the International Congress on Acoustics*, Aachen, Germany, 2019, pp. 5828–5835.

[15] J. L. Verhey and S. J. Heise, "Suprathreshold perception of tonal components in noise under conditions of masking release," *Acta Acustica united with Acustica*, vol. 98, pp. 451–460, 2012. DOI: http://dx.doi.org/10.3813/AAA.918529.

[16] J. Hots, S. A. Vaghefi, and J. L. Verhey, "The effect of sensorineural hearing loss on suprathreshold perception of tonal components in noise," *JASA Express Lett.*, vol. 2(8), pp. 1–6, Aug. 2022. DOI: https://doi.org/10.1121/10.0013352.