

COVER PAGE

Document downloaded by @DAEL

Fri Nov 20 03:57:26 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Quantification of Preference Relevant Sound Characteristics of Multi-Tone Sounds Based on the Differences Between Loudness Judgments and Preference Evaluations, S. Töpken, H. Scheel, J. L. Verhey and R. Weber, Acta Acustica **vol. 104** (Number 1), 2018, pp. 153-165

DOI
https://doi.org/10.3813/AAA.919155

Acta Acustica united with Acustica
Vol. 104 (2018) 153 – 165

DOI 10.3813/AAA.919155

Quantification of Preference Relevant Sound Characteristics of Multi-Tone Sounds Based on the Differences Between Loudness Judgments and Preference Evaluations

S. Töpken[1)], H. Scheel[2)], J. L. Verhey[3)], R. Weber[1)]

[1)] Dept. of Medical Physics and Acoustics, Acoustics Group, Carl von Ossietzky University, Carl-von-Ossietzky-Str. 9-11, 26129 Oldenburg, Germany. stephan.toepken@uni-oldenburg.de

[2)] Airbus Operation GmbH, Kreetslag 10, 21129 Hamburg, Germany.

[3)] Dept. of Experimental Audiology, Otto von Guericke University, Leipziger Str. 44, 39120 Magdeburg, Germany

Summary

The aim of this study is to quantify the impact of the ratio between fundamental frequencies and the contribution of the sound elements on the loudness and preference for intricate multi-tone sounds, consisting of two complex tones and additional combination tones. Preference evaluations and loudness judgments were determined in the form of levels at two points of subjective equality (PSEs) compared to a fixed, common reference sound. The level difference between the two PSEs is attributed to additional perceptual aspects of the multi-tone sounds, which in return become tangible and quantifiable in the form of dB-values. In this paper, the entirety of these additional perceptual aspects, apart from the loudness perception, is subsumed under the term *sound character*. The results show a clear differentiation between loudness judgments and preference evaluations by the participants and stable mean values for the level difference attributed to the sound character were obtained. The results demonstrate that a combination of two complex tones can be beneficial, whereas the addition of the combination tones has no adverse impact on the evaluation of this type of sound. For the evaluation of combined sounds, also the ratio between the fundamental frequencies affects the perception.

PACS no. 43.50á.p, 43.66.Cb, 43.66.Jh, 43.66.Lj

1. Introduction

Tonal components are an emerging part of transportation noise which humans perceive every day. They are especially audible inside vehicles like buses, trains and aircraft, where they usually originate from rotating machinery or engine parts. Prominent tonal components are often unwanted because they can lead to higher annoyance and noisiness, even if the overall sound pressure level is kept constant [1, 2]. A reduction of the pleasantness due to the perception of tonal components occurred for instance in the evaluation of interior sounds in cars [3], trains [4, 5] and aircraft [6]. Especially for turboprop driven aircraft, loudness and tonality were identified to play a role for the annoyance or unpleasantness of the sounds [7]. Västfjäll and co-authors investigated affective reactions to interior aircraft sounds. They identified that variations of the tonal composition affected the pleasantness dimension of the affect space, whereas the activation dimension was related to the overall tone-to-noise ratio for aircraft cabin noise [8, 9]. Mixson *et al.* found that equal comfort between broadband aircraft interior noise and noise containing tonal components may be achieved by reducing the overall level of the latter by 5 dB [6]. However, the tonal signatures investigated so far were mainly single complex tones consisting of one fundamental frequency and higher harmonics.

For future aircraft engine concepts, the tonal signature may be more complex than current signatures. The resulting multi-tone sounds are expected to consist of two complex tones and additional combination tones due to nonlinear aeroacoustic sound generating mechanisms [10, 11]. Such multi-tone sounds substantially differ from single complex tones in terms of perceptual aspects like the psychoacoustic sharpness [12] and roughness [13, 14, 15], both of which are also found to be related to the evaluation of a sound [8, 16, 17, 18]. In a recent study of three of the authors, the perceptual space of such multi-tone sounds, consisting of two complex tones and additional combination tones, was investigated for a variation of the frequency ratio between the two fundamentals [19]. The results of a semantic differential revealed a close link between the

Received 30 November 2016,
accepted 23 December 2017.

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

pleasant factor and the temporal periodicity of the sounds. However, the adjective pairs *pleasant - unpleasant* and *not loud - loud* loaded both highly on the pleasant factor. Due to this correlation between the loudness and the unpleasantness, the quantitative contribution of the loudness on the one hand and the sound character on the other hand to the unpleasantness evaluation is of high interest. Following the dichotomy between loudness and other perceptual aspects, the phrase *sound character* is used hereinafter as an umbrella term for all perceptual aspects influencing the appraisal and evaluation of a sound apart from loudness. The present study focuses on a quantitative differentiation between the contribution of the loudness and the sound character with regard to preference evaluations for specific multi-tone sounds.

In many cases, the evaluation of a sound turns out to be mainly linked to its physical or perceived intensity which often dominates the influence of the sound character [20]. However, the ratio between annoyance evaluations and the loudness sensation is often larger than unity which means that the rate of increase with rising sound pressure level is larger for annoyance than for loudness in these cases. This was found for synthetic sound like tones embedded in noise [21] and also for sounds with a specific sound character [22, 23]. Ratios between annoyance and loudness larger than unity were also observed for aircraft flyover noise [24], for community noises [25, 26] and environmental sounds [27]. Similarly, differences between loudness judgments and annoyance evaluations were also found for synthetic sounds. Niese determined discrepancies between the points of subjective equality (PSEs) for loudness and for annoyance and found level differences of about 4 dB between the two for howling noise [28]. Ronnebaum and Weber found differences between equal loudness and equal unpleasantness contours determined for sinusoids and narrow-band noise of up to 10 dB in the frequency range around 30 Hz [29]. Kurakata *et al.* identified even bigger level differences between equal unpleasantness and equal loudness of up to 20 dB at rather high frequencies around 4 kHz and 10 kHz [30]. All these studies show that, besides high correlation coefficients, there is a perceptual difference between loudness and annoyance which can be quantitatively expressed in the form of level differences. This is also taken into account by the concept of level adjustments which is a part of several international and national standards [31, 32, 33]. Level adjustments are included into the rating level of noise immissions as a correction factor, aiming at a more adequate depiction of the nuisance effects of specific sound characteristics.

The aim of the present study is a separate determination of loudness judgments and preference evaluations as points of subjective equality (PSEs) with respect to a reference noise in the specific application context of aircraft cabin interior noise. The PSEs are quantified in distinct experiments as levels at equal loudness and at equal preference in comparison to one common reference sound. Due to the rather unpleasant sound character of the particular multi-tone sounds investigated here, the PSEs for preference are expected to be lower than the PSEs for loudness as outlined in Figure 1. The resulting level difference between the two PSEs may be interpreted as a level penalty because the level of a multi-tone sound needs to be reduced to attain the transition from equal loudness to equal preference. Accordingly, this level difference provides a quantitative measure for the additional effect of the (unpleasant) sound character to the preference evaluation which comes on top of the perceived loudness of a sound. In the following, it will therefore be denoted as $\Delta L_{\text{sound character}}$.

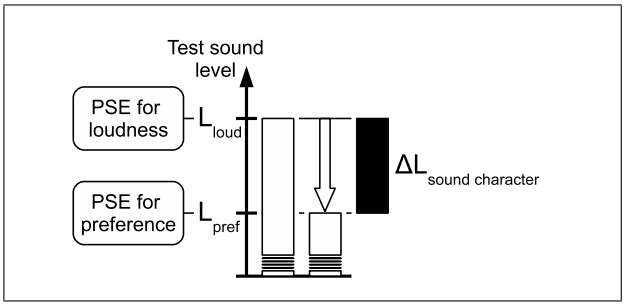


Figure 1. Level scheme: The reference sound is kept constant in level while the multi-tone test sound level is varied by an adaptive procedure on the one hand converging at the PSE for loudness (loudness task, left bar) and on the other hand at the PSE for preference (preference task, middle bar). The resulting levels L_{loud} and L_{pref} represent the points at which the test sound is equally loud and equally preferred as the reference sound (not shown here). The difference between L_{pref} and L_{loud} is attributed to the sound character and thus defined as $\Delta L_{\text{sound character}}$ (right bar).

In the first perception experiment of the present study, five signals varying in the ratio between the fundamentals were investigated with the PSE-method. These signals were selected from a larger set of 15 signals from a previous study with a semantic differential [19]. The five sounds selected for the current study were representatives for the three groups of sounds which had been identified in the previous study by the factorial analysis of the semantic differential data. The five sounds included the most pleasant and the most unpleasant sounds of the previous study of Töpken and co-authors. The aim of this first experiment is a quantification of the loudness and the preference in terms of dB-values, to distinguish between the contributions of the loudness and the sound character to the preference evaluation of these sounds. In the second experiment, the role of the different constituents of the complex multi-tone sounds is investigated. Under the assumption that each sound element contributes its part to the level difference $\Delta L_{\text{sound character}}$, it is expected that $\Delta L_{\text{sound character}}$ is more negative for the combined sounds than for each of the sound constituents. The goal of this experiment is to identify which of the constituents contributes most to the evaluation.

2. Method

In this study, the loudness and the preference of test sounds, are both quantified by the levels at the points of subjective equality (PSEs) in relation to the loudness and

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

preference of a reference noise. Given the observation that a decrease in level reduces the loudness as well as the annoyance or unpleasantness of a sound, the following three steps are taken to quantify preference relevant aspects of the sound character using an adaptive paired comparison procedure:

1. PSE for loudness: Determination of the dB(A)-level L_{loud} of the test sound, at which the test sound is equally loud as the reference sound which is fixed in level.
2. PSE for preference: Determination of the dB(A)-level L_{pref} of the test sound, at which the test sound is equally preferred as the constant level reference sound.
3. Sound character: Quantification of the influence of the sound character on the evaluation by the level difference between both PSE-levels:

$$\Delta L_{\text{sound character}} = L_{\text{pref}} - L_{\text{loud}}. \tag{1}$$

It is assumed that the PSEs for preference L_{pref} can be expressed by the quantitative loudness judgment L_{loud} and an additional term which subsumes all qualitative perceptual aspects related to the sound character which are not already covered by the loudness judgment. This additional term is denoted $\Delta L_{\text{sound character}}$.

$$L_{\text{pref}} = L_{\text{loud}} + \Delta L_{\text{sound character}} \tag{2}$$

In this way, the multidimensionality of the sound character is mapped to the dimension that affects the preference of the sound and merged into a single value with the unit dB. Hence, sound character is defined here as the property that differentiates a test sound which is equally loud as the reference sound from the reference sound in terms of the preference. If no such difference exists, then equal loudness is equivalent to equal preference and

$$\Delta L_{\text{sound character}} = 0 \text{ dB}. \tag{3}$$

On the other hand, if equal preference is not equivalent to equal loudness due to preference relevant differences between the multi-tone test sound and the reference sound in terms of the sound character, then

$$\Delta L_{\text{sound character}} \neq 0 \text{ dB}. \tag{4}$$

The definition of sound character given here is similar to the approach in the studies of Cardozo and co-authors [22, 23, 34]. They aimed at decomposing the acoustic factors contributing to sound annoyance into two independent dimensions. In their view, the first dimension reflected the intensity of the sound in terms of the A-weighted sound pressure level. Their second dimension summarized all other acoustic factors and it was denoted sound character. They further tried to project the observed sound character differences on the L_{eq}-scale "to see how far the L_{eq} is off the mark as a yardstick of annoyance" [23]. In our study, the PSEs for loudness reflect the individual intensity of the sounds instead of the A-weighted levels, which were used by Cardozo and co-authors. Instead of absolute annoyance which was used by Cardozo *et al.*, we asked for preference in comparison to a common reference sound, which can be regarded as an indirect approach to the assessment of sound annoyance, aiming at the more general underlying principles of noise perception [35]. The paired comparison between the test sounds and the constant reference sound used in this study is a rather easy task for the participants with a high degree of reliability. In addition, the continued usage of the common reference noise in each task constitutes at the same time a common acoustic frame of reference.

2.1. Stimuli

In this study, multi-tone sounds which are very rich in sound character differences were selected as test sounds. The digitally generated stimuli were composed as superpositions of two harmonic complex tones CX1 and CX2 with fundamental frequencies f_{10}, f_{01} ($f_{10} < f_{01}$) and additional combination tones CTs with frequency components f_{ij} as basic elements. Similar sounds were used in earlier studies of the authors [19]. In detail, the frequency components of the test stimuli were as follows:

$$\text{1. complex tone CX1:} \quad f_{i0} = i \cdot f_{10}, \quad i = 1, 2, 3, \ldots, 30, \tag{5}$$

$$\text{2. complex tone CX2:} \quad f_{0j} = j \cdot f_{01}, \quad j = 1, 2, 3, \ldots, 30, \tag{6}$$

$$\text{combination tones CTs:} \quad f_{ij} = i \cdot f_{10} + j \cdot f_{01}, \quad i, j = 1, 2, 3, \ldots, 20. \tag{7}$$

Thus, each entire multi-tone test sound including all three elements (CX1+CX2+CTs) consisted of 460 partial tones. All partials had a random starting phase, taken from the same uniform distribution of random numbers between 0 and 2π. The same random phase values were used for the generation of all stimuli to avoid changes in timbre due to different phase information [36]. All partials decreased by 6 dB/octave in level towards high frequencies. The combination tones CTs were attenuated by 10 dB compared to the two complex tones before the superposition of all partials. All stimuli had a duration of 5 seconds in the preference task and 1.5 seconds in the loudness task. The frequency of the lower fundamental f_{10} was fixed for all test conditions ($f_{10} = 100$ Hz). The reference sound was a typical cabin noise signal of a turbofan driven aircraft with a spectral slope of approximately -6 dB/octave up to 1 kHz and -12 dB/octave above 1 kHz. The level of the reference sound $L_{\text{ref}} = 74$ dB(A) was kept constant in the adaptive procedure. This playback level was chosen to reflect typical sound pressure levels of cabin noise during cruise flight [37, 38]. All stimuli were digitally generated and stored in a computer with a sampling rate of 22050 Hz and a resolution of 16 bits.

For a first set of stimuli, the ratio between the fundamental frequencies $\rho = f_{01} : f_{10}$ was varied by changing the frequency f_{01} of the fundamental of the second complex tone CX2 as an experimental parameter for five test sounds, each consisting of all three basic elements

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

(CX1+CX2+CTs). Table I shows the details of both fundamental frequencies f_{01} and f_{10} for all five stimuli. Depending on the value of the frequency ratio ρ between the fundamental frequencies, the superposition of the two complex tones and combination tones can lead to amplifications/cancellations of partials in cases where the frequencies exactly match. In all other cases, beats and modulations emerge due to the closely spaced neighboring partials. Ratios of small integer numbers between the fundamental frequencies (e.g. ρ = 4:3) yield matching partials already at a rather low partial frequency. For ratios of large integer numbers (e.g. ρ = 199:150) the first matching partials did not occur until rather high partial numbers. The values of the ratio ρ included three ratios equal to ratios of small integer numbers (sounds A1, C1 and E1 with ρ = 9:7, 13:10 and 4:3) as well as two stimuli with a slight detuning Δf_{01}, resulting in ratios of rather large integers (sound B1 and D1 with ρ = 1293:1000 and ρ = 199:150). Sound B1 was the detuned version of sound C1 and sound D1 was the detuned version of sound E1. The signals included representatives of the three sound groups found with the semantic differential in the previous study [19]. The most pleasant (ρ = 1293:1000 and ρ = 199:150) and the most unpleasant sound (ρ = 4:3) of the previous study were also included in this stimulus set. Based on the results of the previous study it was hypothesized that the PSEs for loudness and for preference should be lower for the rather unpleasant sounds, which were based on values of ρ equal to ratios of small integer numbers, compared to sound which were built from ratios of large integer numbers.

With a second set of stimuli, the contribution of the basic sound elements to the subjective judgments and evaluations was investigated for a fixed ratio of ρ = 199:150. The five combinations of basic elements are shown in Table II. The complex tones alone D4 and D5 (CX1 and CX2) consisted of 30 partials each. The combination tones D3 (CTs) contained 400 partials. Thus, the superposition of the two complex tones D2 (CX1+CX2) resulted in 60 partials and the entire multi-tone sound with all three elements D1 (CX1+CX2+CTs) consisted of 460 partials. It was hypothesized that each of the elements and in particular the introduction of combination tones to a mixture of the two complex tones should increase the unpleasantness of the multi-tone sounds, due to modulations resulting from a closer spectral spacing of the partials. This augmented unpleasantness would result in lower PSE levels.

2.2. Procedure

In two separate matching experiments, the PSE levels of the test sounds were measured for preference and loudness in comparison to the fixed reference sound. For the determination of the PSEs, the 50 percent points (P50) of the psychometric curves were used. They were measured using a two alternative-forced-choice (AFC) paradigm with a 1-up 1-down (simple staircase) rule [39, 40] by varying the level of the multi-tone test sound. The technique of loudness matching used here is a common procedure of indirect loudness measurements, which is sometimes considered to be the "gold standard" to which results obtained by other methods must conform [41].

Table I. Stimulus set 1: Parameters of the five sounds A1–E1 consisting of all basic elements (CX1+CX2+CTs). Fundamental frequencies f_{10}, f_{01}, detunings Δf_{01} with respect to the corresponding exact ratio of small integer numbers and the resulting ratios of the fundamental frequencies $\rho = f_{01} : f_{10}$. The sound D1 was also the basis for the sound elements of the second stimulus set (Table II).

ID	f_{10} (Hz)	f_{01} (Hz)	Δf_{01} (Hz)	$\rho = f_{01} : f_{10}$
A1	100	128.57	0	9:7
B1	100	129.30	-0.7	1293:1000
C1	100	130.00	0	13:10
D1	100	132.66	-0.66	199:150
E1	100	133.33	0	4:3

Table II. Stimulus set 2: Sound D1 and its combined and individual elements D2-D5 with description.

ID	Elements	Description
D1	CX1+CX2+CTs	Entire multi-tone sound, consisting of booth complex tones and combination tones
D2	CX1+ CX2	Two complex tones, no combination tones
D3	CTs	Combination tones only, no complex tones
D4	CX1	1. complex tone only (f_{10}=100 Hz)
D5	CX2	2. complex tone only (f_{01}=132.66 Hz)

In the preference task, the participants were asked to decide which sound (either the reference or the multi-tone sound) they prefer as a cabin interior noise and preference relevant differences in terms of sound characteristics should also be taken into account by the participants for their preference decision. Depending on the answer of the participants, the level of the multi-tone sound was varied adaptively. Based on the close link between the sound pressure level and the appraisal of a sound, it is assumed that, if the level of the multi-tone sound is considerably attenuated compared to the fixed level of the reference sound, then it will become more pleasant and hence, will be preferred. If the level of the multi-tone sound is set higher than the reference, then the tonal characteristic of the multi-tone sounds are even more pronounced which will lead to a preference decision in favor of the broadband reference sound. Accordingly, the level of a multi-tone test sound is reduced, every time after it is not preferred and raised if it is preferred in the adaptive procedure. The level of the multi-tone test stimulus was initially varied with a step size of 6 dB, which was halved after each upper reversal of the level curve until a minimum step size of

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

1.5 dB was reached. Thereafter, the actual measurement phase started which stopped after six reversals. The two sounds, with a duration of 5 seconds each, were separated by a pause of 1 second. A pause of 2 seconds was introduced between the answer and the presentation of the next stimulus pair.

The measurement of the loudness was carried out using the same paradigm but now asking the question "Which sound is louder?" The level of a multi-tone test sound is decreased every time after it is judged to be louder than the fixed reference sound and the level is increased, if the reference sound is judged to be the louder one, like in classic loudness matching experiments [41]. The step size of the adaptive procedure began at 3 dB and was then halved after the second upper reversal of the level curve down to a step size of 1.5 dB. The tracking stopped after 6 reversals. Between the two sounds - test and reference - each lasting for 1.5 seconds, a pause of half a second was inserted. After another pause of 1 second after the answer of the participant, the next judgment pairs were presented.

In both (preference and loudness) tasks, the presentation order of reference and test sound was randomized within each trial. The level of the reference sound was always kept constant ($L_{\mathrm{ref}} = 74$ dB(A)). At the beginning of each adaptive track, the level of the multi-tone test sound was the same as the reference sound level. During each measurement session, the preference and loudness experiments were carried out for the same sound set consisting of five or six test sounds each. Thus, the preference and the loudness experiments consisted of five or six interleaved adaptive tracks for the five or six test sounds respectively. The resulting individual PSEs are described as levels L_{pref} and L_{loud}, calculated as the mean over six level reversal-points of the adaptive track during the measurement phase (with a step size of 1.5 dB). The average standard deviation from the six reversal points over all participants was 1.5 dB for the loudness PSEs and about 2 dB for the preference PSEs, showing that the decision criterion was generally quit sharp in both tasks. In rare cases this standard deviation was larger, but no systematic structure with respect to specific signals or participants was obvious. Therefore, it was decided to accept a small amount of noise in the individual data and increase the overall number of participants. According to Equation (1), the level difference $\Delta L_{\mathrm{sound\ character}}$ is individually calculated for each participant based on the PSEs of the separate loudness and preference experiments.

All measurements have been carried out within a bigger framework such that the measurements of the data presented in this paper were interleaved with other stimuli not presented here. Overall, four fixed sets of sounds were judged on separate days with about a week in between the experiments. About 25% of the participants started with one of the four sets each, continuing successively with the three other sets. One set consisted of five sounds (sounds A1-E1), the other three sets consisted of six sounds, each. They included the sounds D2-D5, repetitions of the sounds D1 and E1 and 12 other sounds which varied in the amount of combination tones and the shape of the spectral envelope and which are not presented in the frame of this paper. The repetitions of the sounds D1 and E1 were included in the third sound set. This repeated measurement of the two sounds allows for an assessment of the intra-individual reliability and the reproducibility of the mean values between test and retest conditions. Even though a balance over all participants was striven for, the accompanying sounds in the test an the retest condition and the prior experience with other sounds were different for each of the four groups of the participants.

Each experiment was inaugurated by handing out written instructions to the participant. After clarifying still open questions, the first experiment - loudness or preference - took place inside an anechoic chamber. After finishing the first experiment, a short break was introduced before starting the next one. Here also written instructions introduced the participant to the task, which was carried out again inside the anechoic chamber. About one half of the participants started each session with the preference task and the other half performed the loudness task first in each session.

2.3. Apparatus

The experiments were conducted in the anechoic chamber of the University of Oldenburg (lower limiting frequency: 50 Hz). The task itself was implemented as a Matlab (The Mathworks) routine on a computer [42]. An external audio interface (M-Audio, Fast Track Pro) supplied the audio signals to an active loudspeaker (Mackie, HR 824), which was positioned in front of the seated participant inside the anechoic chamber. The distance between the loudspeaker and the listener was 2.25 m. The experimental routine was operated by the participant via a computer keyboard and a TFT-screen, placed underneath the loudspeaker. The setup was calibrated using a handheld sound level meter (B&K 2226) at the listening position above the empty chair. The calibration was checked before the first and after the last experiment each day of the measurement campaign.

2.4. Participants

The experiments were carried out by 47 paid volunteers (23 female, 24 male), mainly students and members of the University of Oldenburg. The mean age of the participants was 23 years (min=19 years, max=31 years). Approximately one half of the participants already participated in previous psychoacoustic experiments (10 female, 13 male), the other half was inexperienced with psychoacoustic tests (13 female, 11 male). All participants reported no hearing difficulties.

3. Results

The intra-individual reliability of the raw loudness and preference results is presented in section 3.1. The influence of the frequency ratio ρ between the fundamental frequencies f_{01} and f_{10} on the level difference $\Delta L_{\mathrm{sound\ character}}$ is

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

given in Section 3.2. The contribution of different multitone sound elements to the level difference $\Delta L_{\text{sound character}}$ is presented in Section 3.3.

3.1. Reliability of the judgments

Repeated measurements were conducted for two test sounds to get an insight into the intra-individual reliability and the reproducibility of the judgments averaged over the participants. The individual results of the 47 participants for sound D1 (ρ = 199:150) and for sound E1 (ρ = 4:3) are shown in Figure 2. Plotted are the individual data of the repeated measurement (retest) over the initial evaluation (test) for the loudness (top row) and the preference task (bottom row). For both sounds the results of the preference task (Figure 2c and d) show considerably more inter-individual spread (in the sense of overall covered range) and also more intra-individual variability (in terms of deviation from the bisecting identity line) than the results of the loudness task (Figure 2a and b). The loudness is judged more reliably by each participant and more uniformly by all 47 participants. In all cases the correlation coefficients between test and re-test are statistically significant with values between $r = 0.8$ ($p < 0.01$) and $r = 0.89$ ($p < 0.01$). Averages of the individual data for sound D1 presented in Figure 2a and 2c are shown in Figure 3a. The averages for sound E1 individually shown in Figure 2b and 2d are plotted in Figure 3b. The results of a statistical test (ANOVA, rep. meas.) are given in Table III.

For the loudness judgments, differences of about 1 dB between the mean values of the test and the retest are statistically significant for both test sounds ($p < 0.01$). This might indicate a potential bias resulting from the repeated measurement, but the magnitude of the differences between test and retest is barely above the just noticeable differences in level (JNDL [18]).

For the preference evaluations, the differences between the mean values of test and retest are about 3 dB and hence makes them statistically significant for both sounds (D1, $p < 0.01$ and E1, $p < 0.05$). A possible explanation for the difference between the mean values are the rather extreme and unreliable evaluations by few participants in the retest condition. An exclusion of those five participants with differences between test and retest larger than 10 dB renders the differences between test and retest insignificant for sound D1 ($F(1, 41) = 2.0$, p=0.169) and also for sound E1 ($F(1, 41) = 1.6$, p=0.216). In addition, the individually different accompanying sounds in the test and retest condition for the four groups of participants and different prior experience with other sound groups judged beforehand cannot be ruled out as a potential source for a bias.

No statistically significant differences were found between the initial and repeated measurement for the level difference $\Delta L_{\text{sound character}}$ (D1, $p = 0.07$ and E1, $p = 0.10$). Even if a systematic bias affected the loudness and preference judgments, it seems as if the level difference is not affected by such a bias. This finding suggest that the calculation of the difference between the two PSEs might be a possible way to control for biases which affect both

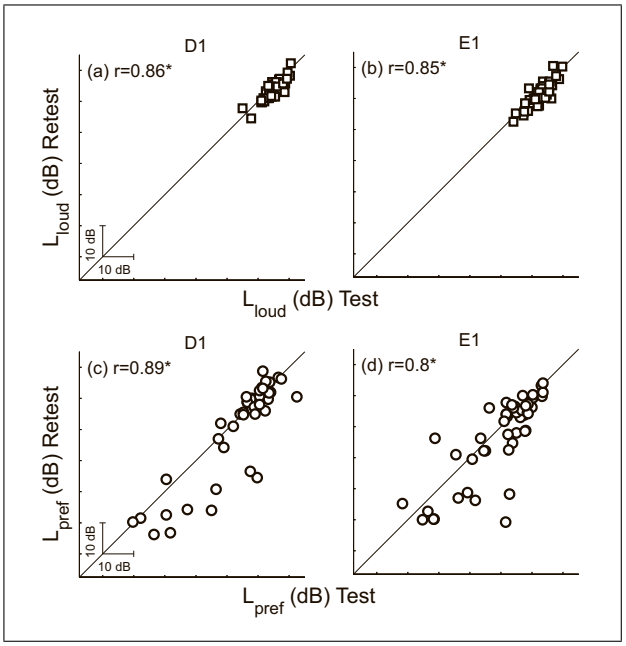


Figure 2. Comparison of test and retest data. Individual results of the repeated measurement (retest) are plotted over the initial judgments (test) for the loudness (a, b) and the preference task (c, d) performed with the sound D1 (left, ρ = 199:150) and sound E1 (right, ρ = 4:3) from the same participant sample (N=47). The bisecting lines indicate identity. A considerable inter-individual variability can be seen for the preference evaluations (c, d), especially at large attenuation values where individuals did not reproduce their evaluations well. The loudness (a, b) is judged more uniformly over the 47 participants, and the judgments are also individually more reliably reproduced. In all cases, the correlation coefficients between test and re-test are statistically significant ($p < 0.01$).

Table III. ANOVA results of the repeated measurements from the test and retest judgments for the sounds $D1$ and $E1$ with N=47 each.

ID		ANOVA results
D1	loudness	$F(1, 46) = 16.3, p < 0.01$
	preference	$F(1, 46) = 7.5, p < 0.01$
	sound character	$F(1, 46) = 3.6, p = 0.07$
E1	loudness	$F(1, 46) = 12.8, p < 0.01$
	preference	$F(1, 46) = 6.2, p < 0.05$
	sound character	$F(1, 46) = 2.9, p = 0.10$

PSEs in a similar way. For the two sounds D1 and E1, which have been repeatedly measured, the mean values based on the data of all 47 participants of the initial test are presented in the following.

3.2. Influence of the ratio between the fundamental frequencies on the judgments

The results of the loudness and the preference evaluations for a variation of the frequency ratio ρ are shown in Figure 4a. A significant main effect of the ratio ρ is found for the loudness judgments ($F(3.280, 150.865) = 31.8, p < 0.01$) and also for the preference evaluations

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

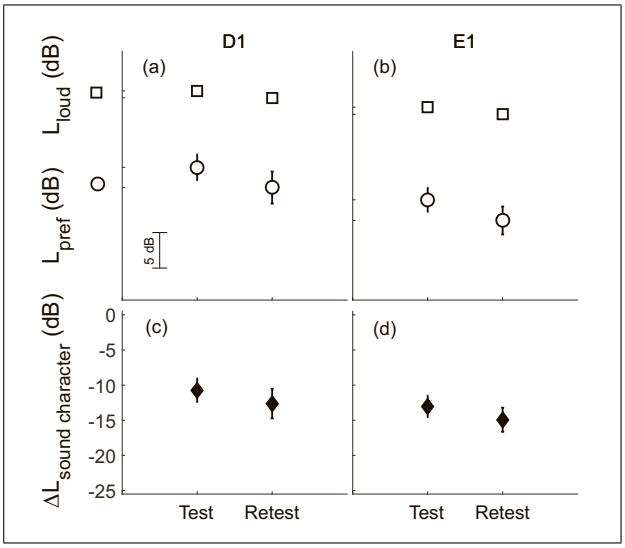


Figure 3. Mean values of the loudness (a, b, open squares) and the preference (a, b, open circles) PSEs and the resulting level difference attributed to the sound character (c, d, filled diamond) for the sound D1 and sound E1 repeatedly measured by the same participant group (N=47). Errorbars indicate standard errors of the calculated mean (sometimes even smaller than the symbols). Ticks on the ordinate in subfigure (a) and (b) reflect the values of the data for a better comparison with the relative scale.

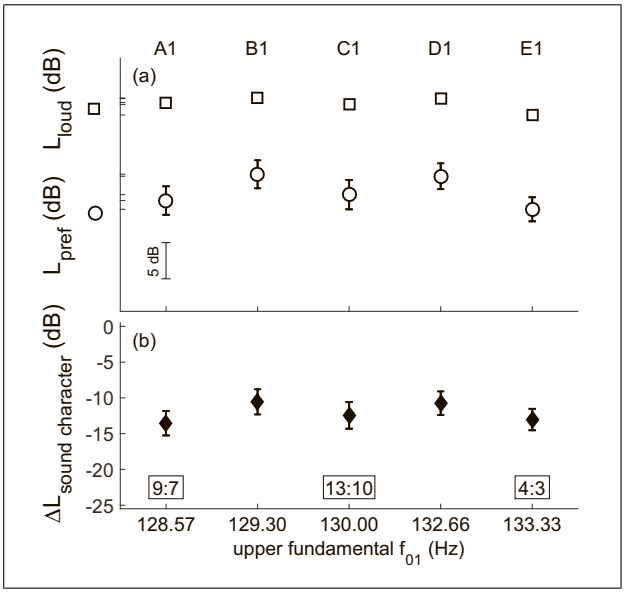


Figure 4. (a) Mean PSE levels for the loudness (open squares) and for the preference (open circles) and (b) the level difference $\Delta L_{\text{sound character}}$ (filled diamonds) plotted for the five ratios of the fundamental frequencies ρ (sound $A1$-$E1$). The resulting upper fundamental frequencies are given at the bottom. Errorbars indicate the standard deviation of the calculated mean (N=47, sometimes even smaller than the symbols). Ticks on the ordinate in subfigure (a) reflect the values of the data for a better comparison with the relative scale.

$(F(2.182, 100.358) = 7.6, p < 0.01)$[1]. Post-hoc tests reveal that ratios $\rho = f_{01} : f_{10}$ equal to ratios of small integer numbers (C1, ρ = 13:10 and E1, ρ = 4:3) require lower levels of the test sounds to be perceived as equally loud and equally preferred as the reference sound than the corresponding ratios of large integers (B1, ρ = 1293:1000 and D1, ρ = 199:150) which are detuned versions of ratios of small integers, each. This means that, at the same A-weighted level, the sound E1 (ρ = 4:3) would be the loudest sound compared to all other four sounds. For the equal loudness judgments the differences in level are about 2 dB between the multi-tone test sounds. In terms of preference, the multi-tone test sounds with frequency ratios ρ based on small integer numbers require up to 5 dB lower levels than values of ρ based on ratios of large integer numbers to become equally preferred as the reference sound.

Figure 4b shows the influence of the ratio ρ on the level difference attributed to the sound character. Values of $\Delta L_{\text{sound character}}$ from −10.5 dB to −13.5 dB are observed. The results of the post-hoc test reveal that the difference of about 3 dB between a ratio of ρ = 9:7 (A1) and 1293:1000 (B1) and the difference of about 2 dB between 1293:1000 (B1) and ρ = 13:10 (C1) are statistically significant ($F(2.4, 108.2) = 3.4, p < 0.05$). The differences between all other pairs were not statistically significant.

3.3. Contribution of the basic sound elements to the judgments

The contribution of the basic sound elements CX1, CX2 and the CTs to multi-tone sounds consisting of superpositions of these basic elements was investigated for a fixed ratio ρ = 199:150 between the fundamental frequencies. The mean values of the loudness judgments and the preference evaluations for the basic sound elements and two combinations of them are shown in Figure 5a.

A significant main effect of the five different elements is found for the loudness judgments ($F(2.217, 101.981) = 57.7, p < 0.01$) and also for the preference evaluations ($F(2.445, 112.455) = 33.8, p < 0.01$). Post-hoc tests on the loudness judgments show significant differences between the single complex tones (D4 and D5, $p < 0.01$) and compared to each of the three other sounds (D1, D2 and D3, $p < 0.01$). The two complex tones alone (D4 and D5) require up to 5 dB lower levels to be equally loud as the combination tones (D3) and the combined sound elements (D1 and D2). The complex tone with the higher fundamental frequency (D5) is significantly different from all other stimuli (D1 - D4) and, thus, also different from the complex tone with the lower fundamental frequency (D4). In the case of the preference evaluations, post-hoc tests reveal that the mean difference between the combinations D1 (CX1+CX2+CTs) and D2 (CX1+CX2) are not statistically significant. Accordingly, it would make no difference for the preference evaluation whether or not the combination tones D3 (CTs) were added to a superposition of the two complex tones D2 (CX1+CX2). However, the basic elements alone (D3, D4 and D5) require considerably

[1] All reported F values result from analyses of variances (ANOVAs) based on repeated measurements and including a Greenhouse-Geisser correction. All presented post-hoc tests include Bonferroni corrections where necessary.

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

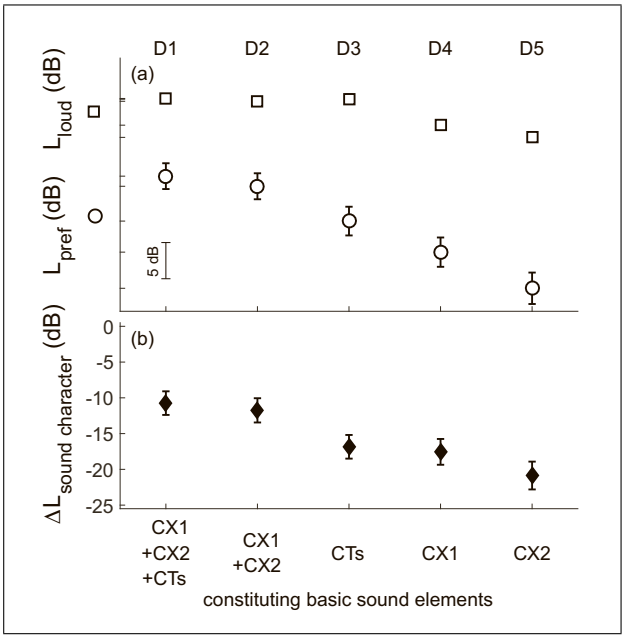


Figure 5. (a) Mean PSE levels for the loudness (open squares) and for the preference (open circles) and (b) the level difference $\Delta L_{\text{sound character}}$ (filled diamonds) plotted for the complete multi-tone sound D1 (CX1+CX2+CTs), the mix of the two complex tones D2 (CX1+CX2), the combination tones only D3 (CTs), and the two complex tones D4 (CX1) and D5 (CX2). Errorbars indicate the standard deviation of the calculated mean (N=47, sometimes even smaller than the symbols). Ticks on the ordinate in subfigure (a) reflect the values of the data for a better comparison with the relative scale. The basic elements need considerably lower levels to become equally loud and equally preferred as the combinations of these elements.

lower levels than the two combined sounds D1 and D2 to be equally preferred. In detail, the complex tone with the higher fundamental frequency D5 (CX2) needs the lowest level compared to the complex tone with the lower fundamental D4 (CX1) and the combinations tones alone D3 (CTs). Therefore the sound D5 would be the least preferred among the sounds of the study if all had the same A-weighted level. The differences between all other pairs were not statistically significant ($p > 0.05$).

The influence of the constituting basic sound elements on the level difference attributed to the sound character is shown in Figure 5b. Values of $\Delta L_{\text{sound character}}$ from -10.7 dB to -20.9 dB reflect the considerable difference between equal preference and equal loudness with respect to the reference sound and also among the five multi-tone test sounds D1, D2, D3, D4 and D5. In both cases, this indicates differences in the sound character. A range of about 10 dB for $\Delta L_{\text{sound character}}$ among the test sounds yields statistical significance ($F(2.7, 122.9) = 17.7, p < 0.05$). The results of a post-hoc paired comparison are similar to those of the raw preference evaluations. For the basic elements alone D3 (CTs), D4 (CX1) and D5 (CX2) the difference in $\Delta L_{\text{sound character}}$ is considerably larger than for the compositions of these elements in the stimuli D1 (CX1+CX2+CTs) and D2 (CX1+CX2) which both do not significantly differ from each other ($p > 0.05$). Out of the three basic elements, the complex tone with the higher fundamental frequency D5 (CX2) is significantly different from D4 (CX1), too ($p < 0.05$).

4. Discussion

4.1. Loudness and preference results

The current findings of the first experiment for the variation of the ratio ρ between the fundamental frequencies are in a very good agreement with the results of a previous study of the authors [19]. The results of that previous study showed that ratios of small integers are rather unpleasant and louder than ratios of large integers. The present study confirms this finding. Sounds based on ratios of small integers (ρ = 9:7, 13:10, 4:3) require more attenuation to become equally preferred and the PSE levels for preference are about 5 dB lower for them than those for the sounds based on ratios of large integers (ρ = 1293:1000, 199:150). The results of the loudness judgments follow a similar structure but the dynamic range of the PSE levels is smaller in this case than for the preference evaluations. The largest differences among the multi-tone sounds at the PSEs for loudness are about 2 dB. In the previous study, a close link between the loudness and the unpleasantness was found and the adjective pairs "not loud-loud" and "pleasant-unpleasant" both loaded highly on the same factor. It was the aim of the present study to quantify the effect of the ratio between the fundamental frequencies on the loudness and the preference with the PSE method to differentiate between the two. The current results now show the quantitative difference between the loudness judgments and the preference evaluations in terms of the level difference $\Delta L_{\text{sound character}}$ in the order of -12 dB. The high correlation coefficient between the loudness judgments and the preference evaluations observed in the current study (Figure 6a) indicates a similar systematic link between the preference evaluations and the loudness judgments as in the previous study which is further discussed in Section 4.2.

The results of the second experiment surprised us. In contrast to our expectation that each sound element contributes a part to the overall PSE levels, the PSE levels for loudness and for preference are considerably lower for the individual sound elements with a low number of partials (sound D4 and D5) than for the combined multi-tone sound (CX1+CX2, sound D2). Thus, the combination of tonal components can even be beneficial in terms of the evaluation of multi-tone sounds. Only about 5 dB of this effect can be attributed to perceived loudness differences. Thus, the remaining 5-10 dBs are related to the sound character according to the definition used in this paper. The regular spacing of the partials in the complex tones with Δf = 100 Hz (sound D4) and Δf = 132.66 Hz (sound D5) results in periodicities that may be linked to modulation frequencies which can be associated to a sensation of roughness. For amplitude-modulated tones with center frequencies above 500 Hz, the maximum roughness is ob-

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

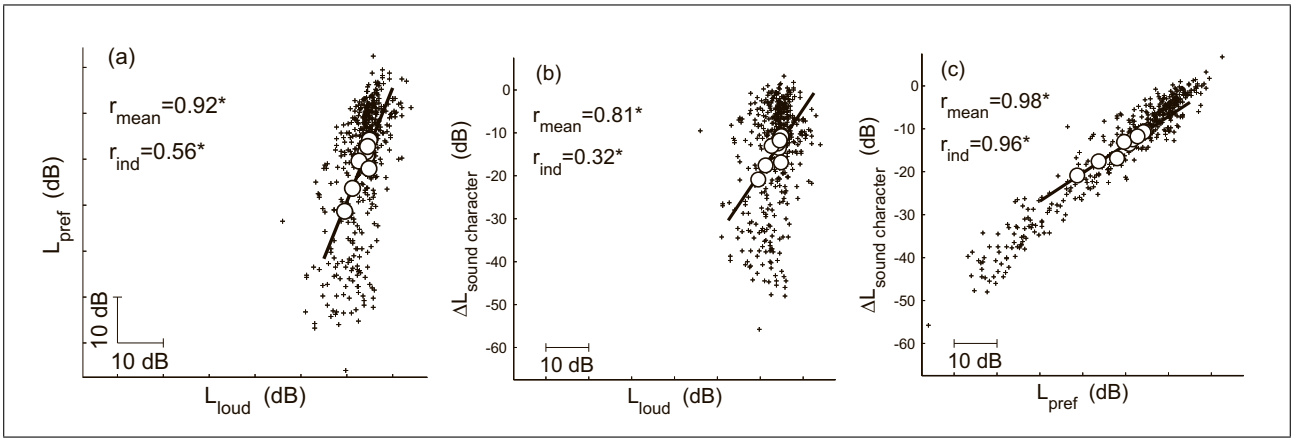


Figure 6. Relationships between the PSEs L_{loud}, L_{pref}, and $\Delta L_{\mathrm{soundcharacter}}$: Individual data (crosses) and mean values (open circles). Linear regressions based on the mean values are indicated as straight lines. (b) A rather low correlation coefficient ($r_{ind} = 0.32$) is observed between the individual data for $\Delta L_{\mathrm{soundcharacter}}$ and L_{loud}. (c) The correlation coefficient between $\Delta L_{\mathrm{soundcharacter}}$ and L_{pref} is very high ($r_{ind} = 0.96$), showing that the additional attenuation $\Delta L_{\mathrm{sound\ character}}$ is closely linked to the preference evaluation.

served for a modulation frequency of 70 Hz and usually, an introduction of periodic fluctuations is detrimental to the pleasantness of a sound [18]. However, the combination of two complex tones with a ratio $\rho = 199{:}150$ will result in rather slow beats between the partials of the two complex tones [15]. In the present study, the beat frequency is not the same for all partials but it increases with the absolute frequency of the higher partials. In addition, the beats in different frequency ranges will not be in phase due to the random phase applied to the partials in the synthesis which may reduce the sensation of beats and roughness considerably [13]. Thus, the comprehensive change of the sound character from a single stationary complex tone producing a pronounced roughness sensation to a mixture of two complex tones with diverse modulations as a function of frequency might be an explanation for the present result.

However, no influence on the preference evaluations was found for the addition of the combination tones (CTs) to the combined two complex tones (CX1+CX2). The data suggests that, for a ratio $\rho = 199{:}150$, the addition of the combination tones has no further adverse impact on the evaluation of the combined multi-tone sound. The loudness judgments are also not affected by an addition of the combination tones. For the mixture of the two complex tones and the combination tones (CX1+CX2+CTs), the level of the combination tones (CTs) was chosen to be 10 dB lower than the overall level of the combined two complex tones (CX1+CX2) to imitate an additional nonlinear sound generation. The data show that the resulting level of the combination tones as a part of the mixture (CX1+CX2+CTs) is about 10 dB below the level at the PSE for loudness and still 5 dB below the level at the PSE for preference which were determined for the combination tones (CTs) alone. This might explain why the addition of the combination tones has no further effect in both cases.

In general, the current findings for the level difference $\Delta L_{\mathrm{sound\ character}}$ are in overall agreement with the results of Ronnebaum and Weber [29], Kurakata *et al.* [30] and Verhey *et al.* [43]. They all found that the level at the PSE for (un-)pleasantness or annoyance is below the level of the PSE for loudness. Their level differences were up to −20 dB between the two PSE levels. In the current study, the additional attenuation between equal loudness and equal preference is described by $\Delta L_{\mathrm{sound\ character}}$. It reaches similar magnitudes of up to −20 dB between equal preference and equal loudness levels for single complex tones (sound D5, CX2, Figure 5) even though the tonal signature is considerably different to those investigated in the studies of Ronnebaum and Weber [29], Kurakata *et al.* [30] and Verhey *et al.* [43]. In the current study, the preclusion of a broadband background from the multitone test sounds was chosen as an exaggeration of a sound with very prominent tonal sounds which was intended to cover a worst case estimate. As such, a purely tonal sound without any broadband sound is unlikely to occur in a flying aircraft and the considerable differences expressed as $\Delta L_{\mathrm{sound\ character}}$ might be interpreted as an indication that purely tonal sounds are considerably less acceptable in the context of aircraft cabin noise than the common aircraft noise used as the reference [44, 45].

The choice of the reference sound is of course a key factor for the loudness and preference matches. In a previous study of some of the authors, loudness and preference matches were determined for pink noise containing tonal components using white noise and pink noise as reference sounds [46, 47]. In these studies, the preference PSEs were unaffected by the choice of the reference. In contrats to this, the loudness PSEs were all shifted by about 4 dB for the white noise reference, but the relative differences between the test sounds did not change. An explanation for this shift was the loudness difference between pink noise and white noise for a fixed A-weighted sound pressure level. For the level difference $\Delta L_{\mathrm{sound\ character}}$ introduced in this paper, such an effect of the reference sound would alter the absolute values of $\Delta L_{\mathrm{sound\ character}}$ but the relative differences would be expected to remain stable. Accordingly, also the observed correlation coefficients between the PSEs illustrated in Figure6 would be expected to re-

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

main similar even for a different reference sound. For the application context of aircraft cabin noise addressed in this study, the choice of a rather neutral cabin noise signal as the reference sound seems to be appropriate.

The inter-individual differences between the 47 participants are larger for the preference evaluations than for the loudness judgments in the current study. This is in agreement with the findings of Berglund *et al.* [26]. The inter-individual standard deviations of the mean loudness judgments over the 47 participants from the current study are about 3 dB i.e., a standard error of 0.4 dB. This standard deviation is comparable with literature data on loudness matches [48, 49].

4.2. Relationship between loudness, preference and sound character judgments

Figure 6a shows the relationship between the preference evaluations and the loudness judgments as a scatter plot of the individual data as well as the mean data of all nine investigated sounds on aligned relative scales. The quantitative differences between the PSEs for loudness and the PSEs for preference show that the participants required different levels to make the multi-tone sounds equally loud on the one hand and to make them equally preferred on the other. The values of the individual data for L_{pref} are generally lower than L_{loud}, meaning that additional attenuation compared to the equally loud level is necessary to make the multi-tone test sounds equally preferred as the reference sound. This difference in the results for the two tasks indicates that the participants could indeed distinguish between the loudness and the preference task. However, a correlation coefficient of $r_{\mathrm{ind}} = 0.56$ ($p < 0.01$) is found between the L_{loud} and the L_{pref} for the individual data and $r_{\mathrm{mean}} = 0.92$ ($p < 0.01$) for the mean values shows a systematic dependency between loudness and preference. A regression line in Figure 6a indicates the relationship between the mean values of the preference and the loudness judgments for the nine sounds. Interestingly, the value for the regression slope of 2.44 (Figure 6a) is similar to the relationship found by Hiramatsu *et al.* for environmental sounds [27]. In their study a regression slope of 3.11 was found between the levels of the PSEs for annoyance and loudness calculated indirectly from categorial scaling for a wide variety of everyday environmental sounds and white noise of different levels[2].

The current findings are also in agreement with literature data where high correlation coefficients between loudness and unpleasantness [8, 50] or between loudness and annoyance judgments [25, 27, 51] were found, especially from laboratory experiments including a level variation of the sounds under test [52].

The level difference between the two PSE levels for loudness and preference, which was initially defined as $\Delta L_{\mathrm{sound\ character}}$ provides a more detailed insight into the difference between the loudness and preference judgments. The relationship between $\Delta L_{\mathrm{sound\ character}}$ and the original loudness judgments L_{loud} is displayed in Figure 6b. A rather low correlation coefficient of $r_{ind} = 0.32$ ($p < 0.01$) between the individual data of $\Delta L_{\mathrm{sound\ character}}$ and L_{loud} is found. A resulting shared variance of only about 10 percent stresses the rather weak relationship between the individual data of loudness and sound character. This weak link suggests that the level difference, which was initially defined as $\Delta L_{\mathrm{sound\ character}}$, is not linearly related to the loudness sensation but reflects a different aspect. Additionally, Figure 6c reveals a close relationship between $\Delta L_{\mathrm{sound\ character}}$ and the PSEs for preference L_{pref} with 92 percent of shared variance. These three relationships indicate that the participants may reliably state preference evaluations in the form of level differences compared to a fixed reference sound, which are clearly distinguished from the loudness judgments. The findings show that the criterion for preference can be disassembled into a part due to loudness and into a part due to the sound character. The measurement method employed here allows to quantitatively determine both constituents of the preference by introducing $\Delta L_{\mathrm{sound\ character}}$. The level difference $\Delta L_{\mathrm{sound\ character}}$ can be regarded as the preferential difference between sounds having the same loudness.

4.3. Relationship between the judgments and the repetition rates of the signals

In a previous study with similar multi-tone sounds, high correlation coefficients between the repetition rate of the signals and unpleasantness judgments determined with a semantic differential and a paired comparison were found [19]. In that study, the repetition rate was defined as the inverse of the periodicity of the time signals. The periodicity can be derived from the lag of the first unity peak of the autocorrelation function or it can be calculated as the respective integer fraction from the fundamental frequencies[3]. A variation of the ratio ρ between the two fundamental frequencies directly affects the periodicity and the repetition rate of a multi-tone signal. Values of ρ equal to ratios of small integers (e.g. $\rho = 4{:}3$) lead to short periodicities and high repetition rates and ratios of large integer numbers (e.g. $\rho = 199{:}150$) yield long periodicities and low repetition rates. In the previous study, multi-tone sounds based on values of ρ equal to ratios of small integer numbers were judged more discordant and unpleasant than sounds based on ratios of large integer values. This finding is in contrast to classic musical consonance theory [53, 54] but in good agreement with studies dealing with multi-tone sounds consisting of a high number on partial tones which are precisely controlled in frequency [15, 55, 56].

Figure 7 shows the results of the current study as a function of the repetition rate f_{rep}. Linear regressions over the results of all nine signals, indicated as solid lines in Figure 7, yield an adjusted coefficient of determination

[2] In Figure 3 of the Hiramatsu paper the assignment of regression equations to the subplots is obviously wrong. The bottom left plot should be denoted (b) instead of (c) and vice versa.

[3] Further details on the calculation and a table with values of the repetition rate for the signals can be found in the reference [19].

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

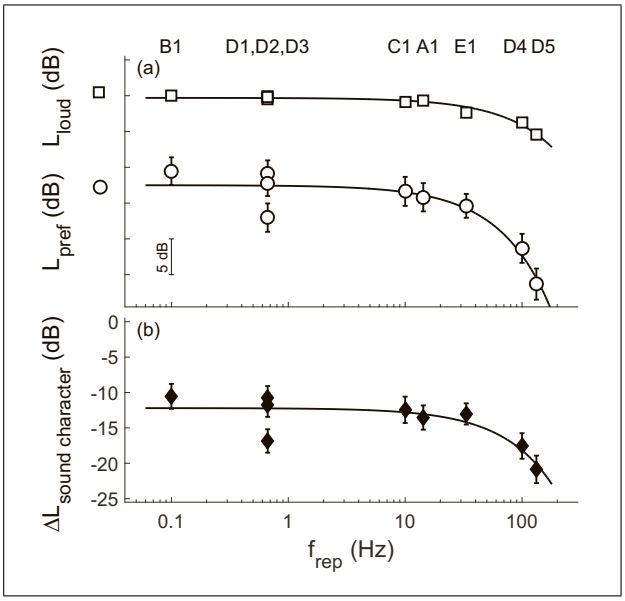


Figure 7. (a) Mean PSE levels for the loudness (open squares) and for the preference (open circles) and (b) the level difference $\Delta L_{\text{sound character}}$ (filled diamonds) plotted for all nine signals as a function of their repetition rate f_{rep}. Solid lines indicate the results of linear regressions.

$r^2_{\text{adj}} = 0.96$ for L_{loud}, $r^2_{\text{adj}} = 0.85$ for L_{pref} and $r^2_{\text{adj}} = 0.66$ for $\Delta L_{\text{sound character}}$. Except for sound D3 which consists of the combination tones only, the line of the linear regression progresses within the standard errors of the measured data. An explanation for the deviation of sound D3 from the linear regression line for the preference PSEs and for $\Delta L_{\text{sound character}}$ might be a difference in sharpness. Sound D3 has a sharpness of 1.4 acum according to the DIN 45692 standard whereas all other signals are in the range from 0.9 to 1.1 acum [57]. For the sounds with similar sharpness values, the results suggests that low repetition rates (which are equivalent to long periodicities) are preferred over high repetition rates (which are equivalent to short periodicities) which is in good agreement with the results of the previous study [19]. One explanation for this finding might be a link between the repetition rate and the perceived roughness of the multi-tone sounds. The spectral structure of the complex tones alone (sound D4 and D5) fulfills two major requirements for a pronounced roughness sensation. On the one hand, the partials are spectrally near enough to each other such that the resulting periodicities are very close to the modulation frequency for which roughness is maximal [18, 58]. On the other hand, they are sufficiently spaced apart from each other in frequency that a summation of the partial roughness in the lower auditory bands could be presumed [13, 14]. The sounds consisting of all three elements (CX1+CX2+CTs) have a rather high number of very closely spaced partials in each critical band and an interaction of multiple neighboring partials for the resulting beats is inevitable. Due to the different random starting phase of each of the partials, the resulting beats will also have random phases which presumably include antiphasic beatings that may reduce the roughness for these sounds [13, 14]. Furthermore, lower roughness would also be expected for the sounds with lower periodicity values (ratios of large integers) than for the sounds with higher periodicity values (frequency ratios of small integers) [15, 56]. A more detailed investigation of the relationship between the repetition rate of a signal and the produced roughness could be a meaningful extension of the present study. It is worth noting that the rather high coefficients of determination cannot be solely traced back to the high values of the repetition rate for sound D4 and D5 which are about a magnitude higher than the rest of the signals. A linear regression over the results from sound A1 to E1 (variation of the frequency ratio ρ) only still leads to $r^2_{\text{adj}} = 0.95$ for L_{loud}, $r^2_{\text{adj}} = 0.86$ for L_{pref} and $r^2_{\text{adj}} = 0.59$ for $\Delta L_{\text{sound character}}$.

The relationship between the repetition rate and the loudness suggests a decrease of the loudness PSEs for increasing repetition rates. This means that multi-tone sounds with higher repetition rates are louder than those with lower repetition rates. An effect of about 5 dB between a repetition rate of 10 Hz and 100 Hz is observed. This is in general agreement with the data of Zwicker and Fastl, but they found a steeper increase in loudness level of 10 phon for an increase of the repetition rate by factor of 10 for tone bursts [18]. However, the current finding is in contrast to literature data indicating a louder sensation for two- and three-tone sounds producing beats [18, 59]. The multi-tone signals of the present study consist of a considerably higher number of partials than the two- or three-tone experiments in the literature studies making a direct comparison to this data difficult. Due to the close spacing of the higher partials of the combination tones in the present study, slow beats between the fundamental frequencies will occur together with a variety of beat frequencies for the higher partials which might result in a different sensation than the beats or amplitude modulations from two or three tones in the mentioned studies. Even though the absolute values of the PSEs for loudness and for preference differ considerably, a tight correlation between the loudness and the preference judgments is generally observed (Figure 6(a)). Therefore, an influence of the preference task on the loudness judgments can not be completely ruled out and also the within-subject design might have contributed to this tight link. Further investigations on the influence of specific modulation on the loudness might be a part of future work.

5. Conclusion

The aim of this study was to reliably distinguish between the loudness and the preference of multi-tone sounds in the context of aircraft cabin noise. Preference was chosen instead of annoyance. Multi-tone sounds were compared to a reference sound and the PSE for loudness and preference were determined by adapting the levels of the multi-tone sounds. The level difference between the PSE level for preference and the PSE level for loudness is attributed to the sound character: $\Delta L_{\text{sound character}} = L_{\text{pref}} - L_{\text{loud}}$.

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

The results show that a precise quantitative determination of the sound character in the form of a level difference between loudness judgments and preference evaluations was successful. The findings can be summarized as follows:

1. For all nine investigated multi-tone sounds $\Delta L_{\text{sound character}}$ was negative which means that equal loudness was not sufficient to make the multi-tone sounds also equally preferred as the reference sound. Additional level reductions were necessary to achieve equal preference with the reference sound.
2. Frequency ratios $\rho = f_{01} : f_{10}$ between the two fundamentals that could be expressed as ratios of small integers (e.g. ρ = 4:3) required more negative values of $\Delta L_{\text{sound character}}$ (hence greater level reductions) than frequency ratios of large integer numbers (e.g. ρ = 199:150 $\approx$ 4:3) which is in qualitative agreement with a previous study [19]. The quantitative effect was about 5 dB between the preference PSEs L_{pref} and about 3 dB for $\Delta L_{\text{sound character}}$.
3. In contrast to the initial expectation that each sound element contributes its part to the unpleasantness of the multi-tone sounds, the combined sounds (CX1+CX2 and CX1+CX2+CTs) allow for up to 15 dB higher levels at the PSE for preference than the individual sound elements (CX1, CX2 and CTs) for a ratio between the fundamentals of ρ = 199:150. Thus, the combined sounds have been found to be considerably more pleasant than the individual elements when having the same A-weighted level. About 5 dB of this effect can be attributed to perceived loudness differences.
4. For a ratio between the fundamentals of ρ = 199:150, the combination tones, when added to the two superimposed complex tones, did not significantly change the loudness judgments and the preference evaluations of the two superimposed complex tones alone. This was valid for the case when the overall level of the combination tones was 10 dB lower than the overall level of the combined two complex tones. Hence, a moderate amount of combination tones resulting from non-linear sound generating processes seems to have no significant impact on the preference evaluation for this type of multi-tone sounds.

Acknowledgement

The authors thank Airbus Operation, who financially supported the study. Parts of the data presented in this paper were presented at Inter-Noise 2011 [60] and Inter-Noise 2012 [61].

References

[1] R. P. Hellman: Loudness, annoyance, and noisiness produced by single-tone-noise complexes. J. Acoust. Soc. Am. **72** (1982) 62–73.

[2] R. P. Hellman:. Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness. J. Acoust. Soc. Am. **77** (1985) 1497–1504.

[3] H. Hansen, R. Weber, U. Letens: Quantifying tonal phenomena in interior car sounds. Proc. Forum Acusticum 2005, Budapest, 2005, 1725–1729.

[4] C. Patsouras, H. Fastl, U. Widmann, G. Hölzl: Psychoacoustic evaluation of tonal components in view of sound quality design for high-speed train interior noise. Acoust. Sci. Tech. **23** (2002) 113–116.

[5] E. Parizet, N. Hamzaoui, J. Jacquemoud: Noise assessment in a high-speed train. Appl. Acoust. **63** (2002) 1109–1124.

[6] J. S. Mixson, G. C. Greene, T. K. Dempsey: Sources, control, and effects of noise from aircraft propellers and rotors. Technical Memorandum TM-81971, NASA, Langley Research Center, 1981.

[7] J. R. Angerer, D. A. McCurdy, R. A. Erickson: Development of an annoyance model based upon elementary auditory sensations for steady-state aircraft interior noise containing tonal components. Technical Memorandum 104147, NASA, Langley Research Center, 1991.

[8] D. Västfjäll, M. Kleiner, T. Garling: Affective reactions to interior aircraft sounds. Acta Acust. United Ac. **89** (2003) 693–701.

[9] D. Västfjäll: Emotional reactions to tonal and noise components of environmental sounds. Psychology **4** (2013) 1051–1058.

[10] D. B. Hanson: Noise of counter-rotation propellers. J. Aircraft **22** (1985) 609–617.

[11] D. E. Van Zante, E. Envia: Prediction of the aero-acoustic performance of open rotors. Technical Report NASA/TM-2014-218132, NASA, Glenn Research Center, 2014.

[12] G. von Bismarck: Timbre of steady sounds: A factorial investigation of its verbal attributes. Acustica **30** (1974) 146–159.

[13] W. Aures Ein Berechnungsverfahren der Rauigkeit (A procedure for calculating auditory roughness). Acustica **58** (1985) 268–281.

[14] P. Daniel, R. Weber: Psychoacoustical roughness: Implementation of an optimized model. Acta Acust. United Ac. **83** (1997) 113–123.

[15] A. Miskiewicz, T. Rogala, J. Szepanska-Antosik: Perceived roughness of two simoultaneous harmonic complex tones. Arch. Acoust. **32** (2007) 737–748.

[16] W. Aures: Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen (The sensory euphony as a function of auditory sensations). Acustica **58** (1985) 282–290.

[17] A. Preis: Measurement of annoyance components. In F. Alison Hil and Roy Lawrence, editors, Proc. Inter-Noise 1996, Liverpool. Institute of Acoustics, 1996, 5:2289–2294.

[18] E. Zwicker, H. Fastl: Psychoacoustics: Facts and Models. 3rd ed. Springer, Berlin/Heidelberg, 2007.

[19] S. Töpken, J. L. Verhey, R. Weber: Perceptual space, pleasantness and periodicity of multi-tone sounds. J. Acoust. Soc. Am. **138** (2015) 288–298.

[20] C. Marquis-Favre, E. Premat, D. Aubree: Noise and its effects: A review on qualitative aspects of sound. Part II: Noise and annoyance. Acta Acust. United Ac. **91** (2005) 626–642.

[21] R. P. Hellman: Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum. J. Acoust. Soc. Am. **75** (1984) 209–218.

[22] B. L. Cardozo, R. A. J. M. van Lieshout: Estimates of annoyance of sounds of different character. Appl. Acoust. **14** (1981) 323–329.

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only

[23] B. L. Cardozo, R. A. J. M. van Lieshout: Splitting sound annoyance into two components: level and sound character. Annual Progress Report 16, IPO - Instituut voor Perceptie Onderzoek - Institute for Perception Research, 1981.

[24] B. Berglund, U. Berglund, T. Lindvall: Scaling loudness, noisiness, and annoyance of aircraft noise. J. Acoust. Soc. Am. **57** (1975) 930–934.

[25] B. Berglund, U. Berglund, T. Lindvall: Scaling loudness, noisiness, and annoyance of community noises. J. Acoust. Soc. Am. **60** (1976) 1119–1125.

[26] B. Berglund, A. Preis, K. Rankin: Relationship between loudness and annoyance for ten community sounds. Env. Int. **16** (1990) 523–531.

[27] K. Hiramatsu, K. Takagi, T. Yamamoto: A rating scale experiment on loudness, noisiness and annoyance of environmental sounds. J. Sound Vib. **127** (1988) 467–473.

[28] H. Niese: Beitrag zur Relation zwischen Lautstärke und Lästigkeit von Geräuschen (On the relationship between loudness and nuisance of sounds). Acustica **15** (1965) 236–243.

[29] T. Ronnebaum, R. Weber: Lautstärke und Unangenehmheit von Schallen im tieffrequentem Bereich (Loudness and unpleasantness of low frequency sounds). Fortschritte der Akustik - DAGA 1996, Bonn, Germany. DEGA e.V., Berlin, 1996, 350–351.

[30] K. Kurakata, T. Mizunami, K. Matsushita: Sensory unpleasantness of high-frequency sounds. Acoust. Sci. Tech. **34** (2013) 26–33.

[31] ISO1996-1:2003: Acoustics - Description, measurement and assessment of environmental noise - Part 1: Basic quantities and assessment procedures. International Organisation for Standardisation, 2003, 1–32.

[32] DIN 45645-1:1996: Determination of rating levels from measurement data - Part 1: Noise immission in the neighbourhood. Beuth, Berlin, Germany, German industry standard, 1996, 1–9.

[33] DIN 45645-2:1997: Determination of rating levels from measurement data - Part 2: Noise immissions at the workplace. Beuth, Berlin, Germany, German industry standard, 1997, 1–12.

[34] B. L. Cardozo, K. G. van der Veen: Estimation of annoyance due to low level sound. Applied Acoustics **12** (1979) 389–396.

[35] B. Berglund, P. Hassmén, A. Preis: Annoyance and spectral contrast are cues for similarity and preference of sounds. J. Sound Vib. **250** (2002) 53–64.

[36] R. Plomp, H. J. M. Steeneken: Effect of phase on the timbre of complex tones. J. Acoust. Soc. Am. **46** (1969) 409–421.

[37] S. Pennig, J. Quehl, V. Rolny: Effects of aircraft cabin noise on passenger comfort. Ergonomics **55** (2012) 1252-1265.

[38] H. Scheel, P. Lempereur: Next generation Aircraft - A Challenge for Interior Acoustics Developments. Proc. Inter-Noise 2016, Hamburg. DEGA e.V., Germany, 2016, 3254-3263.

[39] H. Levitt: Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. **49** (1971) 467–477.

[40] T. S. Meese: Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. Percept. Psychophys. **57** (1995) 267–281.

[41] L. E. Marks, M. Florentine: Springer Handbook of Auditory Research, volume 37 - Loudness. Springer Science+Business Media, LLC, 2011.

[42] S. D. Ewert: AFC - a modular framework for running psychoacoustic experiments and computational perception models. In AIA-DAGA 2013, International Conference on Acoustics, Merano. DEGA e.V., Berlin, 2013, 1326–1329.

[43] J. L. Verhey, J. Hots, T. Luft, S. Ringwelski, U. Gabbert: A psychoacoustical evaluation of active and passive methods for noise reduction in automotive engineering. AIA-DAGA 2013, International Conference on Acoustics, Merano. DEGA e.V., Berlin, 2013, 1772–1775.

[44] R. Guski: Psychological methods for evaluating sound quality and assessing acoustic information. Acta Acust. United Ac. **83** (1997) 765–774.

[45] J. Blauert, U. Jekosch: Sound-Quality Evaluation - A Multi-Layered Problem. Acta Acust. United Ac. **83** (1997) 747–753.

[46] S. Töpken, S. van de Par, R. Weber: Suitable reference sounds for loudness and preference matching experiments. In Fortschritte der Akustik - DAGA 2015, Nürnberg, Germany. DEGA e.V., Berlin, 2015, 309-312.

[47] S. Töpken, S. van de Par, R. Weber: Subjective evaluation of loudness and preference for noise containing audible tones. J. Acoust. Soc. Am. **137** (2015) 320.

[48] I. Silva, M. Florentine: Effect of adaptive psychophysical procedure on loudness matches. J. Acoust. Soc. Am. **120** (2006) 2124–2131.

[49] W. Heeren, J. Rennies, J. L. Verhey: Spectral loudness summation of nonsimultaneous tone pulses. J. Acoust. Soc. Am. **130** (2011) 3905–3915.

[50] A. Zeitler, J. Hellbrück: Sound quality assessment of everyday-noises by means of psychophysical scaling. Proc. Inter-Noise 1999, Ft. Lauderdale. Stewart Glegg, Yan Yong (editors). INCE, USA, 1999, 1–6.

[51] S. Kuwano, H. Fastl, S. Namba: Loudness, annoyance and unpleasantness of amplitude modulated sounds. Proc. Inter-Noise 1999, Ft. Lauderdale. Stewart Glegg, Yan Yong (editors). INCE, USA, 1999, 1–6.

[52] R. Guski, H.-G. Bosshardt: Gibt es eine "unbeeinfluSSte" Lästigkeit? - does "unbiased annoyance" exist? Z. Lärmbekämpfung **39** (1992) 67–74.

[53] J. M. Geary: Consonance and dissonance of pairs of inharmonic sounds. J. Acoust. Soc. Am. **67** (1980) 1785–1789.

[54] W. A. Sethares: Local consonance and the relationship between timbre and scale. J. Acoust. Soc. Am. **94** (1993) 1218–1228.

[55] H. Fletcher, E. D. Blackham, R. Stratton: Quality of piano tones. J. Acoust. Soc. Am. **34** (1962) 749–761.

[56] V. Vencovsky, F. Rund: Roughness of two simultaneous harmonic complex tones on just-tempered and equal-tempered scales. Music Perception **35** (2017) 127–143.

[57] DIN 45692:2009: Measurement technique for the simulation of the auditory sensation of sharpness. Beuth, Berlin, Germany, German industry standard, 2009, 1–14.

[58] A. Miskiewicz, A. Rakowski, T. Rociszewska: Perceived roughness of two simultaneous pure tones. Acta Acust. United Ac. **92** (2006) 331–336.

[59] C. Zhang, F.-G.Zeng: Loudness of dynamic stimuli in acoustic and electric hearing. J. Acoust. Soc. Am. **102** (1997) 2925–2934.

[60] S. Töpken, J. L. Verhey, R. Weber: Evaluation of multi-tone sounds. Proc. Inter-Noise 2011, Osaka. INCE, USA, 2011, 1–7.

[61] S. Töpken, J. L. Verhey, R. Weber: Preference and loudness of multi-tone sounds. Proc. Inter-Noise 2012/ASME/NCAD, New York. INCE, USA, 2012, 1–8.

Document downloaded by @DAEL, 2026-11-20 03:57:29 - personal use only