Comparing Sonification Strategies Applied to Musical and Non-Musical Signals for Auditory Guidance Purposes

Kantan, Prithvi Ravi

Published in:
Proceedings of the 19th Sound and Music Computing Conference, June 5-12th, 2022, Saint-Étienne (France)

DOI (link to publication from Publisher):
10.5281/zenodo.6573218

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Kantan, P. R. (2022). Comparing Sonification Strategies Applied to Musical and Non-Musical Signals for Auditory Guidance Purposes. In R. Michon, L. Potter, & Y. Orlarey (Eds.), *Proceedings of the 19th Sound and Music Computing Conference, June 5-12th, 2022, Saint-Étienne (France): SMC/JIM/IFC 2022* (pp. 279-286). Sound and Music Computing Network. Proceedings of the Sound and Music Computing Conference https://doi.org/10.5281/zenodo.6573218

Proceedings of the 19th Sound and Music Computing Conference, June 5-12th, 2022, Saint-Étienne (France)

Comparing Sonification Strategies Applied to Musical and Non-Musical Signals for Auditory Guidance Purposes

Prithvi Ravi Kantan
Aalborg University, Copenhagen
`prka@create.aau.dk`

ABSTRACT

Researchers have increasingly explored the potential of interactive auditory guidance in navigation and spatial orientation tasks. Despite laboratory promise, the adoption of these applications in real-life remains limited, partly due to the lack of aesthetic considerations in the design of auditory guidance stimuli, leading to auditory fatigue and low user acceptance. Although music has been suggested as a solution and tested in motor learning and rehabilitation, there is a lack of empirical research comparing its guidance efficacy with traditional nonmusical designs. Through a one-dimensional guidance task with 18 participants, the present study compared an array of novel musical strategies with nonmusical strategies based on the same auditory perceptual dimensions. It was observed that the musical strategies elicited higher user experience ratings while affording comparable performance (error, acquisition time, overshoots) to the nonmusical strategies. There were also performance differences based on the auditory dimensions manipulated by the strategies (e.g. pitch, loudness). There is thus preliminary evidence that music warrants more serious consideration as a means to address the issues of pleasantness and user preference in auditory guidance.

1. INTRODUCTION

Aided by technological advances in the sound and music computing field, digital audio applications have been increasingly explored as assistive tools in otherwise non-audio domains [1]. One such tool is *auditory guidance (AG)*, wherein a human user is interactively guided through a real-life task by an auditory stimulus. Although most conventional guidance systems communicate with human users through the visual modality (e.g. maps, medical displays, traffic signals, etc.), there are certain situations where the auditory modality may be more appropriate. This could be due to the visual modality being absent due to blindness or preoccupied with other tasks, coupled with the fact that the auditory perceptual channel is typically open and capable of processing parallel streams of information [1,2]. The utility of AG has been researched for indoor navigation [3], locomotion guidance [4], pedestrian navigation [5], spatial orientation tasks [6], precision tasks [7], and aircraft control [8] with promising results that demonstrate performance benefits from AG.

An AG signal may be verbal (such as GPS-based navigation instructions on Google Maps) or nonverbal (commonly seen in parking assistance systems), and continuous or discrete. Regardless of the specific application, AG systems work by computing the instantaneous state of the user, comparing it to the state value corresponding to task completion (target state), and using their difference (in the form of a normalized distance) as a control variable for real-time audio synthesis or processing [9]. The output is an audio signal that varies based on this distance in such a way that it guides the user towards the target. This can be seen as a case of interactive sonification, where humans interact with a system that converts data to sound [10]. The sonification topology may be 1D (only one state variable is sonified, e.g. angular distance during indoor navigation [3]) or multidimensional (e.g. two sonified Cartesian coordinates to guide 2D spatial orientation tasks [6]). This can be seen as real-time *parameter mapping sonification* [11], where discrete variables are mapped to separate auditory perceptual dimensions (e.g. pitch, loudness, tempo, timbre), ideally selected based on the task needs [12] as well as the quantity being represented [13]. Many algorithms can be used, from simple methods like subtractive/FM synthesis to more complex ones such as physical models and granular synthesis, or even DSP applied to pre-recorded audio signals such as music tracks [11]. Design rationales may be metaphorical, psychoacoustic, spatialization-based, or abstract [4, 6, 7, 9].

Despite their potential, AG applications enjoy only limited adoption and acceptance in real-life contexts at present [2, 9], an issue that also plagues the field of auditory display and sonification at large [14–16]. Researchers have attributed this to several possible causes that ultimately relate to the lack of interdisciplinary collaboration in the research field [1, 2]; these range from design philosophies and aesthetics to empirical evaluation practices. Specifically, the sonification design process has tended to lack formalization [9], leading to the prevalence of ad-hoc design practices and trial-and-error based research [1]. Design approaches have also tended to be simple and devoid of psychoacoustically or psychologically driven motivations for sound design decisions [2]. Moreover, there is a shortage of studies comparing multiple sonification strategies [2, 9]. Even those that have done so [12] have stuck to simple sound designs and not explored the impact of more aesthetic approaches (e.g. music) on guidance task outcomes.

Here, my objective was to experimentally compare nonmusical and musical AG strategies based on common auditory perceptual dimensions in terms of task performance and user experience in a 1D guidance context. The following sections present related research as well as the design and implementation of the present sonification strategies, which were finally evaluated in a 1D task similar to [12] with 18 normal-hearing participants.

2. RELATED RESEARCH

2.1 Strategies Used in Auditory Guidance

Sonification strategies in AG have been classified based on their perceptual characteristics. Parseihian et al. (2015) [9] classified strategies depending on the presence of a sonic reference denoting the target state. They defined three types - *without reference* (e.g. sine pitch), *with reference* (e.g. synchronicity, harmonicity), and *reference-and-zoom* (e.g. multiband frequency modulation). At a lower level, strategies can also be classified on the basis of the auditory properties they manipulate. Pitch and loudness are commonly used dimensions [9], but more complex psychoacoustic dimensions such as chroma, beating, and roughness have successfully been tested as well [6].

The appropriate sonification strategy for an AG task depends upon the spatiotemporal constraints of the task, and past studies have typically used evaluation metrics such as target acquisition error, time, and overshoots [2,3,6,12]. In short, most studies have found AG to be feasible and beneficial [2–6]. Many have compared the efficacies of AG and visual guidance, mostly finding that visual guidance is faster and more accurate [5, 6], although AG outperforms it in some cases [7]. The former is not surprising, considering that vision is far more precise, and subject to far less individual variability than audition [15]. But even purely within AG, studies have tended to lack rigorous empirical comparisons of different strategy types [2]. A notable exception is [12], where multiple strategies were compared in the context of a 1D target-finding task. The authors found systematic differences in task performance depending on the nature of the sonification strategies (auditory dimension, presence of sonic reference) and the task goal (speed, accuracy, or overshoot minimization). They did not, however, investigate the implications of sonic aesthetics in the context of their designed strategies [12].

2.2 The Issue of Aesthetics

Aesthetics, which binds together aspects such as intrusiveness, listener fatigue, annoyance, and comprehensibility of a sonification [17], can impact the utility, usability, and eventual acceptance of an application. Human-computer interaction research has shown that aesthetic values are an important determinant of user preferences [18], and similar considerations apply to sonification design [2, 9, 19–21]. A lack of aesthetically pleasing and preference-tailored sonification designs may contribute to ongoing problems with acceptance and adoption [2, 9, 20]. Despite the notion that the functional and aesthetic properties of auditory displays cannot be dealt with independently [17, 21], it is common to make use of simple sounds (e.g. sine waves, noise) and perceptual dimensions (qualities - e.g. pitch, loudness) [13]. This, aside from not catering to user tastes, can lead to auditory fatigue [2, 9, 19], although acceptance and fatigue have been found to be subjective [6].

To alleviate these concerns, several approaches have been proposed and tested - these are based on the premise that it is possible to apply variations of an auditory dimension to not only simple sounds, but also complex sound textures without affecting sonification performance [12]. For instance, morphological earcons [9] essentially convey information through *sonic evolution along auditory dimensions*, rather than the chosen sounds themselves; this creates a common 'semantic language'. It is thus possible to satisfy individual preferences and allow seamless switching between various sound palettes without significant changes in cognitive load or learning time [9]. Indeed, many complex 'sound palettes' have successfully been tested, such as physically modeled instruments [7] and natural/electronic soundscapes [9], but we restrict our focus to the one type of organized sound with universal appeal - music.

2.3 Music as an AG Medium

Although musical sonification has been explored to a considerable extent in the domain of motor learning and movement rehabilitation for its ability to motivate and modify movement [22–24], music as a *guidance* medium has been less common. This may be due to the reluctance of designers to use a complex and dynamic signal for fear of the seeming ambiguity introduced to data when it is represented using aesthetic approaches [20]. But while it is indeed true that musical sonification designs have a less direct data-sound relationship (indexicality [17]) than more traditional designs, there is a little empirical evidence that musical AG is prohibitively imprecise as a result. In fact, experiments have shown that adaptive DSP (panning, tempo changes) applied to user-selected music can serve as an effective guidance tool [8,23], and even simple scale melodies have been shown to be effective AG tools [3].

If an AG signal must be continuously audible (as asserted in several works [6, 8]), then music may be a viable means to address the problems of aesthetics and personal preference. The goal of the present study was to compare musical and nonmusical strategies in terms of objective outcomes (e.g. error, speed) and elicited user experience. This could serve as empirical evidence to justify the incorporation of layered musical material in AG design, providing a better understanding of *how* musical materials can be manipulated so as to generate effective AG. In this study, the term 'musical sonification' refers to the use of audio DSP processing on continuous music signals to convey guidance information.

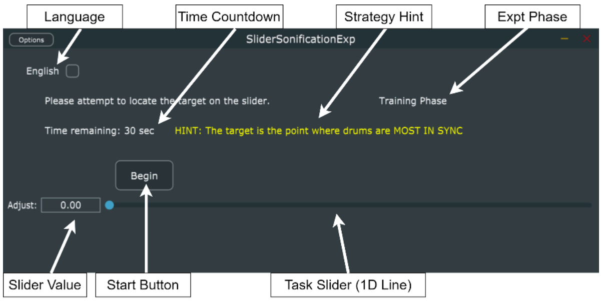
3. MATERIALS AND METHODS

For the present study, two sets of AG strategies were tested - *musical* and *nonmusical*. For the nonmusical set, a subset of the strategies used in [12] were reproduced. The novel musical strategies were designed such that each nonmusical strategy had a musical *correlate* based on the same auditory dimension. Whilst the nonmusical strategy manipulated a stationary signal (sine tone, noise, beep train), its musical counterpart manipulated the characteristics of a synthesized multitrack musical arrangement (described next). The *reference-and-zoom* strategies in [12] were excluded due to the complexity of adapting them for a layered musical signal. The experimental evaluation was a simplified recreation of the 1D target-finding task in [12], where an unspecified point on a line (target) had to be located with the help of AG. To generate the present strategies, real-time music sequencing was carried out in a JUCE-based[1] C++ program, and FAUST[2] was used for audio synthesis, processing, and mixing.

3.1 Music Generation

The sequencing functionality in JUCE was written to provide note (triggering, pitch, velocity) information to the synthesis engine at sixteenth note intervals at 120 BPM. The synthesis engine in FAUST generated eight tracks corresponding to pop/electronic instruments. These included percussive and melodic instruments in several frequency registers, enabling syncopation and true polyphony in a 4/4 time signature. The synthesis methods were based on simple waveforms (e.g. square, sawtooth, noise) with envelope-controlled filters, as well as FM with simple attack-release envelopes. Parametric equalizers and dynamic range compressors are applied so as to attain the desired balance prior to stereo mixdown. The source code can be found in the attached web repository[3]. Melodic and harmonic patterns corresponding to four popular songs[4] were encoded into the sequencer for the purpose of this study.

3.2 AG Strategy Design



Please attempt to locate the target on the slider.

Training Phase

Time remaining: 30 sec

HINT: The target is the point where drums are MOST IN SYNC

Figure 1. The JUCE-built interface used to carry out 1D AG tasks using the different strategies. The mouse was used to initiate tasks and adjust the slider value (user position along the line segment). Tasks were concluded by hitting the space bar.

The strategies were designed for 1D guidance, wherein the task goal was to navigate along a line segment to an unknown *target* point using AG as shown in Fig. 1. Hence the sonified quantity was the distance x between the instantaneous position of the user along the segment and the target. x was normalized between 0 and 1, where 1 corresponded to the total length of the segment. Six auditory dimensions were considered, each with nonmusical and musical AG strategy correlates. The nonmusical strategies are only briefly described here (see [12] for the motivation behind the parameter ranges). Also, the term *melody instruments* refers henceforth to the combination of all pitched instruments (bass synth, chord synths, main melody synth). Audiovisual examples of all strategies are provided[5].

3.2.1 Nonmusical Strategies

- **Pitch (P):** x was positively mapped to the fundamental frequency of a pure tone through a logarithmic function that scaled the tone frequency in the 300 Hz - 3.4 kHz range.

- **Loudness (L):** x was positively mapped through a logarithmic power function to the amplitude of a sine wave at 600 Hz, such that the dynamic range corresponding to the x value range was 40 dB.

- **Brightness (Br):** x controlled the bandwidth of a white noise signal via the cutoff frequency of a 2nd order lowpass filter, such that the cutoff frequency range was 300 Hz - 3.4 kHz (maximum at $x = 0$, with logarithmic scaling).

- **Amplitude Modulation (AM):** A 200 Hz sine tone was added to a second sine tone whose frequency was controlled by x, such that their summation produced a beating/amplitude modulation effect at frequencies ranging from 10 Hz to 0 Hz (no beating at $x = 0$).

- **Synchronicity (Sy):** x was mapped to the temporal offset between two identical beep sequences at the same tempo (120 BPM), such that the beeps were perfectly synchronized when x was 0, and offset by 125 ms when x was 1.

- **Harmonicity (Hm):** x was mapped to the higher partial frequencies of a sinusoidal harmonic series, such that increasing x introduced an inharmonicity factor of up to 0.01.

3.2.2 Musical Strategies

- **Pitch (P):** x was mapped to the root note frequency of the melody instruments as a multiplicative factor, causing them to be transposed upward by a maximum of 60 semitones as:

$$fTonic_{final} = fTonic_{original} \cdot (1 + 5 \cdot x)$$

---

[1] JUCE Framework - `https://juce.com/`

[2] FAUST Language - `https://faust.grame.fr/`

[3] `https://github.com/prithviKantanAAU/SMC2022_AG_Code_Data`

[4] The Eagles - Hotel California, Tom Petty and the Heartbreakers - Free Fallin', Luis Fonsi - Despacito

[5] AG Strategy Demo Video: `https://www.youtube.com/watch?v=RMbzq1rydrY`

- **Loudness (L):** x was mapped to the gain control of the melody instruments, such that increases in x resulted in these instruments becoming softer as:

$$MelGain_{dB} = -80 \cdot x$$

- **Brightness (Br):** x controlled the cutoff frequency of a resonant low pass filter on the melody instruments such that the cutoff frequency was maximum on the target and decreased with increases in x. The filter had a Q-factor of 2.5 to highlight cutoff frequency changes.

$$fc_{Mel}(Hz) = 12000 - 11800 \cdot x$$

- **Amplitude Modulation (AM):** x was mapped to the depth control of a tremolo effect (amplitude modulation) that processed the melody instruments. The tremolo effect consisted of a sine *LFO (low frequency oscillator)* at the musical beat frequency. Increasing x increased the effect intensity as:

$$AMProd = LFO \cdot MelMix_{pre}$$

$$MelMix_{final} = \\ MelMix_{pre} \cdot (1 - x) + AMProd \cdot x$$

- **Synchronicity (Sy):** x was mapped to temporal offset factors separately applied to all instruments (realized using delay buffers), reducing their level of musical synchronization as x increased. The delay in seconds for each instrument was calculated as:

$$d(Inst) = d_{Max}(Inst) \cdot x$$

- **Harmonicity (Hm):** x was mapped to the mix ratio control of a ring modulator effect connected to the melody instruments. The modulation frequency of the effect was configured to the tritone relative to the tonic, as this was found to create the most inharmonic result compared to other frequencies. Increasing x made the output more dissonant as:

$$RMProd = MelMix_{pre} \cdot sin(2\pi f_{mod} t)$$

$$MelMix_{final} = \\ MelMix_{pre} \cdot (1 - x) + RMProd \cdot x$$

4. EXPERIMENTAL EVALUATION

An experiment was carried out with the purpose of directly comparing the AG strategies in terms of guidance efficacy and user experience. At the outset, I framed the following hypotheses based on past literature:

- H1: Participants will exhibit inferior performance with the musical strategies than with their nonmusical correlates, manifesting as greater error, longer acquisition time, and more overshoots per trial.
- H2: There will be differences in task performance between the auditory dimensions (e.g. brightness and synchronicity will lead to greater error than pitch and loudness).
- H3: The musical strategies will be rated as more pleasant and preferable for longer use durations than their nonmusical correlates.

4.0.1 Participants

A convenience sample of 18 participants (3 women) from Aalborg University aged 29.4 ± 6.3 years volunteered to participate in the experiment. Each of them was briefed about the purpose and length of the experiment, and informed that they could withdraw at any time. All experimental procedures conformed to the ethics code of the Declaration of Helsinki. Informed consent was obtained prior to participation, and no sensitive or confidential information was collected from the participants.

4.0.2 Experimental Setup

The experiment was conducted in a quiet room at Aalborg University. It took place on a Dell laptop loaded with the experimental interface (see Fig. 1), which provided instructions, set up the AG tasks, and recorded participant responses. An external keyboard and mouse were used to operate the software, and sound was played using Audio Technica ATH-M50X headphones.

4.0.3 Procedure

After the initial briefing, the participant was asked to complete an online musical background questionnaire to determine their *Ollen Musical Sophistication Index (OMSI)* [25]. The experiment was conducted using the interface, which randomized the presentation order of the 12 AG strategies. For each strategy, the procedure was as follows:

- Training Phase: The participant familiarized themselves with the AG strategy during this phase. An on-screen hint was first provided (see Fig. 1 for an example). Based on this, they attempted to locate a randomly determined target point using the AG strategy. There was no time limit for this, and a feedback prompt appeared when the participant was in the vicinity of the target value. When the participant felt comfortable with the strategy, they could proceed to the main testing phase.
- Testing Phase: This was identical to the training phase, except that there was a 30 second time limit, and no prompt appeared in the target vicinity. The instruction was to find the target point *'as quickly and accurately as possible within the time limit'*. Participants pressed the space bar to conclude the trial when complete, at which time the trial data were logged. Following the guidance task, the participant was given accuracy feedback and asked to rate two qualitative measures on a 7-point Likert scale - (A) how pleasant they found the guidance strategy, and (B) the duration for which they would prefer to use the AG strategy.

The above procedure was repeated for all 12 AG strategies. During both phases, the random target location was constrained to lie between 50% and 85% of the slider length so as to make the initial approach clearly audible and prevent unavoidable overshoots.

4.0.4 Data Analysis

The data were organized in MATLAB 2018b and statistically analyzed in SPSS 27.0. The outcome measures *(abbreviations in parentheses)* were:

- Absolute Error % *(Error)*
- Acquisition Time *(Time)*
- Target Overshoots *(Overshoots)*
- Rated Pleasantness *(Pleasantness)*
- Rated Duration of Preferred Use *(PrefDuration)*

These outcomes were compared both *between and within auditory dimensions*, respectively using repeated-measures (RM) ANCOVA analyses between (i) the six auditory dimensions (each with their two correlated strategies averaged) **[Hypothesis H2]**, and (ii) the 12 AG strategies, with participant OMSI score *(OMSI)* as the covariate. I then examined differences between the musical and nonmusical correlates within each auditory dimension to test **H1** and **H3**. A significance criterion $\alpha = 0.05$ was used for all analyses, and post-hoc Tukey tests were carried out when significant main effects and/or interactions were detected. In addition, Pearson correlation coefficients were computed between each pair of outcomes to investigate possible linear associations. Summary measures for the data are presented as **mean ± standard deviation**. The JUCE code, data logs, analysis scripts, and SPSS test outputs are available on GitHub.

5. RESULTS

We found that the participants (OMSI 252.11 ± 275.06) were able to perform the AG tasks with an overall error of **2.59 ± 1.94 %**, taking **16.17 ± 4.29 sec** to complete each task, and committing **7.16 ± 2.91 target overshoots** per task. Table 1 shows the Pearson correlation coeffients between each pair of outcomes across each of the 216 individual trials (18 participants × 12 trials), revealing several significant correlations. For the ANCOVA analyses carried out across auditory dimensions and strategies, there were no significant main effects of the covariate (OMSI), and between-subject statistics are therefore not reported.

5.1 Between Auditory Dimensions

The RM ANCOVA revealed a significant main effect of *Auditory Dimension* on *Error* (F(2.34, 37.51) = 8.926, p < 0.001, $\eta_p^2 = 0.258$) with post-hoc comparisons revealing significant pairwise differences among several auditory dimensions (Fig. 2 (A)). It is seen that the **Loudness** strategies elicited the least error (0.90 ± 1.36 %) and the **Synchronity** strategies elicited the greatest (5.47 ± 4.79 %).

	Er	**Tm**	**Ov**	**Ple**	**PDur**
Er	1	0.132	**-0.153**	-0.126	-0.067
Tm	-	1	**0.296**	**0.139**	**0.154**
Ov	-	-	1	0.018	-.002
Ple	-	-	-	1	**0.739**
PDur	-	-	-	-	1

Table 1. The upper triangular portion of the Pearson's correlation matrix between all outcomes across the 216 recorded trials. **Bold and underlined** values indicate significant correlations. Er = Error, Tm = Time, Ov = Overshoots, Ple = Pleasantness, PDur = PrefDuration

There was also a significant main effect on *Time* (F(5, 80) = 11.395, $p < 0.001$, $\eta_p^2 = 0.416$), with post-hoc comparisons showing several pairwise differences (Fig. 2 (B)). Participants were fastest when using the **Loudness** strategies (12.48 ± 4.38 sec), and slowest when using the **Amplitude Modulation** strategies (20.40 ± 5.64 sec). There was no main effect of *Auditory Dimension* on *Overshoots* (F(2.83, 45.38) = 1.836, p = 0.157, $\eta_p^2 = 0.103$) or *Pleasantness* (F(5, 80) = 1.895, p = 0.106, $\eta_p^2 = 0.106$). There was, however, a main effect on *PrefDuration* (F(5, 80) = 2.625, p = 0.03, $\eta_p^2 = 0.141$), with post-hoc comparisons revealing significant pairwise differences (Fig. 2 (C)). The **Loudness** strategies received the highest ratings out of 7 (3.94 ± 1.74), whereas the **Harmonicity** strategies received the lowest (2.67 ± 1.31).

5.2 Within Auditory Dimensions

The RM ANCOVA showed a significant main effect of *Strategy* on *Error* (F(3.70, 62.91) = 4.544, p = 0.003, $\eta_p^2 = 0.211$) and *Overshoots* (F(4.34, 73.78) = 2.549, p = 0.042, $\eta_p^2 = 0.13$). Post-hoc comparisons did not, however, show significant differences between correlated strategies within any auditory dimension for either outcome. There was a significant main effect on *Time* (F(11, 187) = 10.678, $p <$ 0.001, $\eta_p^2 = 0.386$). Post-hoc comparisons showed that participants took significantly longer when using musical strategies than their nonmusical correlates for three of the six auditory dimensions **(Pitch, Loudness, Harmonicity)** (Fig. 3 (A)). But on the whole, the qualitative outcomes appeared to favor the musical strategies. There was a significant main effect of *Strategy* on *Pleasantness* (F(11, 187) = 9.279, $p < 0.001$, $\eta_p^2 = 0.353$), with post-hoc comparisons showing that musical strategies were rated significantly higher than their nonmusical correlates for all auditory dimensions except **Synchronicity** (Fig. 3 (B)). Lastly, there was also a significant main effect on *PrefDuration* (F(11, 187) = 4.803, $p < 0.001$, $\eta_p^2 = 0.22$). Post-hoc comparisons revealed higher ratings for the musical strategies corresponding to three auditory dimensions **(Loudness, Brightness, Harmonicity)** (Fig. 3 (C)).

6. DISCUSSION

In this study, I experimentally compared nonmusical auditory guidance strategies [12] with a set of novel musical

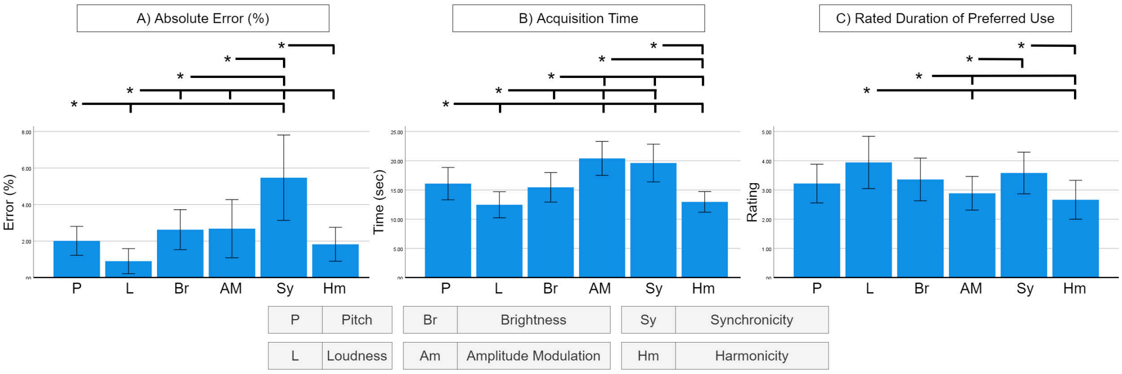


Figure 2. Selected outcomes across auditory dimensions, with constituent musical/nonmusical strategy outcomes averaged. The brackets indicate significant differences between auditory dimensions. Bar heights represent mean values, and error bars show 95% confidence intervals.

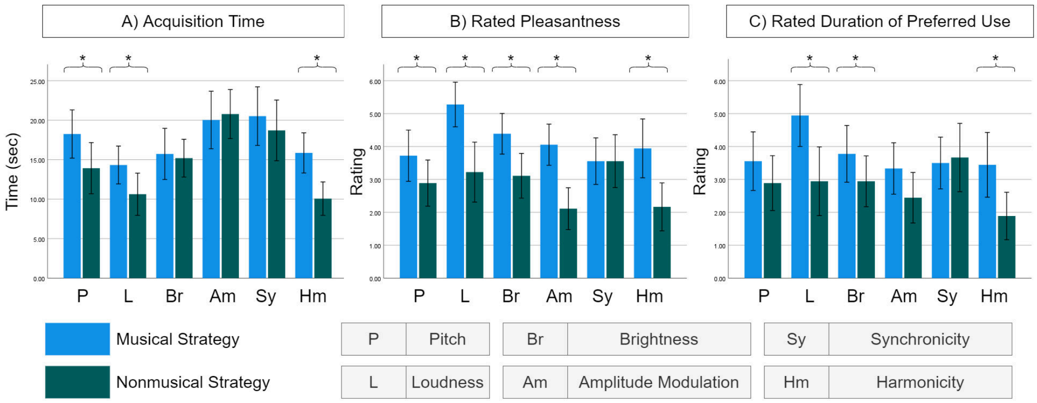


Figure 3. Selected outcomes across the individual AG strategies, with correlated musical/nonmusical strategies stacked together within the respective auditory dimension. The asterisks indicate significant differences between correlated strategies. Bar heights represent mean values, and error bars show 95% confidence intervals.

strategies based on the same auditory dimensions. They were assessed in terms of guidance efficacy and elicited user experience. Based on the known trade-off between aesthetics and precision in data representation through sound [15, 17, 20], I first hypothesized that participants would exhibit inferior performance with the musical strategies than with the nonmusical ones (**H1**). This was partially validated in terms of acquisition time, where participants took significantly longer with musical AG for three of the six auditory dimensions. This can be attributed to heightened uncertainty when examining the pitch, loudness, and harmonicity of a complex, nonstationary signal (music) as compared to simple tones, which may have caused users to take longer to finalize their judgments.

On the other hand, there were no differences between the correlated strategies in terms of error and overshoots. This can be seen as evidence of the musical and nonmusical strategies showing comparable efficacy despite the overall background of the participants (mean OMSI 252.11, meaning 'less musically trained' [25]). But it is also likely that these results were due to the design of the experiment. Unlike in other comparable studies [6, 12] the participants received only limited practice with each AG strategy (1 round of target-finding), making it unlikely that they had reached their performance ceiling before the testing phase. It is therefore premature to comment on the peak performance allowed by the musical and nonmusical strategies.

Furthermore, the experiment did not examine the effect of different instruction types prioritizing speed, accuracy, and overshoot minimization, although past research has shown that this strongly impacts AG user behavior and performance, particularly overshoots [12]. The instruction to find the target 'quickly and accurately' may have meant that individual participants prioritized speed and accuracy differently, making it difficult to replicate the clear differences across strategies seen in [12]. Nevertheless, the significant negative correlation between overshoots and error aligns with past findings [12] in that when prioritizing accuracy (minimizing error), participants carry out small back-and-forth adjustments close to the target that manifest as overshoots. Ultimately, the relatively small sample size (18 participants) and single testing phase trial per strategy (as opposed to three in [12]) may also have contributed to making the differences in error and overshoots between musical and nonmusical strategies harder to statistically detect.

Next, it was hypothesized that participants would perform differently with the various auditory dimensions **(H2)**. This was validated by significant differences in both error and acquisition time. Pitch, loudness, and harmonicity were the best-performing dimensions in these terms, with synchronicity performing poorest in terms of error. In the case of musical AG, this may have been due to difficulties in understanding small deviations in musical timing among several interacting rhythms, especially considering that most participants lacked musical expertise. In terms of time, AM and synchronicity took longest, most likely due to the low frequency of the AM (slow changes in volume), and the relatively large minimum time required to make rhythm-based judgments compared to e.g. pitch [26]. The overall dimensionwise trends in [12] could not be replicated, purportedly due to differences in experiment design.

The final hypothesis **(H3)** was that the musical strategies would be rated more favorably than their nonmusical counterparts in terms of pleasantness and preferred duration of use. This was largely validated by significant differences seen within almost all auditory dimensions. This is in line with past findings about users preferring algorithmic musical sonification over other designs [27], but it is worth noting that this was despite the extremely basic DSP methods used to synthesize the music. It is likely that the ratings for musical AG would be even better with user-selected pre-recorded music tracks like in [8]; most of the musical strategies (except synchronicity) can easily be adapted into DSP algorithms for recorded audio sequences.

There was also a strong correlation between *Pleasantness* and *PrefDuration*, which can be interpreted as a ‘halo-effect’, where the perception of aesthetic values may have strongly influenced overall user preference [18]. Although [18] found *usability* to be the other key predictor of preference, we did not directly capture any subjective measures of usability, and future studies should incorporate these so as to study the usability factors that dictate overall preferences. The user ratings were also positively correlated to acquisition time, either because the participants wilfully spent more time listening to their preferred AG strategies, or because they simply took longer to find the target using musical AG. It is worth noting the task was relatively short (30 sec) and user ratings would probably have changed with longer exposure to all AG strategies [19]. In the absence of significant correlations between the subjective ratings and error/overshoots, it also remains unclear precisely how user preferences are linked to task performance (a.k.a. guidance efficacy). It is likely that this will relationship will vary depending on the needs of the AG task at hand [21].

Limitations of the study include the limited ecological validity of the 1D AG task, static nature of the target value/state, and the lack of equivalence between how musical and nonmusical correlate strategies were generated (e.g. AM and harmonicity). It is unclear whether the findings will generalize to higher-dimensional and/or dynamic tasks, especially considering the increased cognitive load and strict need for orthogonal auditory dimensions in these situations [6, 12, 15]. In terms of gauging user experience, more established questionnaires such as the system usability scale and other HCI standards could have been used, as well as the NASA-TLX for assessing subjective cognitive load. A measure of invasiveness of the sounds was also lacking. Limitations of the experiment design such as instruction type, practice length, and total test repetitions should be addressed in future studies, which should also be conducted with a larger and more gender-balanced sample, preferably with potential users (e.g. visually impaired). Lastly, the OMSI [25] may not have been the best measure of musical ability in this context as it did not reveal any differences depending on music training, something that sonification studies have consistently found [15]. More targeted tests of rhythmic and melodic discrimination may be worth adopting in future investigations.

7. CONCLUSIONS

This study extended past research on auditory guidance by comparing the efficacy of musical strategies with previously established nonmusical ones. Aside from providing an AG performance comparison of the various auditory dimensions when used with a dynamic musical stimulus, the findings suggest that the use of music has the potential to improve user experience when using AG without introducing significant performance degradation. Although future studies should address the limitations of the present methodology and perform rigorous comparisons in real-world AG scenarios, this study provides preliminary evidence that the use of music warrants more serious consideration as a means to address present issues of pleasantness and user preference in auditory guidance design.

Acknowledgments

I would like to thank Sofia Dahl for her valuable inputs during the conception and execution of the study, as well as the individuals who took part in the experiment. This work is partially funded by by NordForsk’s Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

8. REFERENCES

[1] G. Kramer *et al.*, “The sonification report: Status of the field and research agenda. report prepared for the national science foundation by members of the international community for auditory display,” 01 1999.

[2] D. Black, C. Hansen, A. Nabavi, R. Kikinis, and H. Hahn, “A survey of auditory display in image-guided interventions,” *International journal of computer assisted radiology and surgery*, vol. 12, no. 10, pp. 1665–1676, 2017.

[3] D. Ahmetovic, U. Oh, S. Mascetti, and C. Asakawa, “Turn right: Analysis of rotation errors in turn-by-turn

navigation for individuals with visual impairments," in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 2018, pp. 333–339.

[4] F. Dollack, M. Perusquía-Hernández, H. Kadone, and K. Suzuki, "Auditory locomotion guidance system for spatial localization," in *2019 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. IEEE, 2019, pp. 1–5.

[5] A. Montuwy, B. Cahour, and A. Dommes, "Visual, auditory and haptic navigation feedbacks among older pedestrians," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2017, pp. 1–8.

[6] T. Ziemer and H. Schultheis, "Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks," *Journal on Multimodal User Interfaces*, vol. 13, no. 3, pp. 205–218, 2019.

[7] H. Roodaki, N. Navab, A. Eslami, C. Stapleton, and N. Navab, "Sonifeye: Sonification of Visual Information Using Physical Modeling Sound Synthesis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 11, pp. 2366–2371, 2017.

[8] B. D. Simpson, D. S. Brungart, R. C. Dallman, R. J. Yasky, and G. D. Romigh, "Flying by ear: Blind flight with a music-based artificial horizon," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2008, pp. 6–10.

[9] G. Parseihian, S. Ystad, M. Aramaki, and R. Kronland-Martinet, "The process of sonification design for guidance tasks," *J Mob Med*, vol. 9, no. 2, 2015.

[10] A. Hunt and T. Hermann, "Interactive sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Verlag Berlin, 2011.

[11] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. Logos Verlag Berlin, 2011.

[12] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 674–686, 2016.

[13] G. Dubus and R. Bresin, "A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities," *PloS one*, vol. 8, no. 12, p. e82491, 2013.

[14] M. Ballora, "Sonification, science and popular music: In search of the 'wow'," *Organised Sound*, vol. 19, no. 1, pp. 30–40, 2014.

[15] J. G. Neuhoff, "Is Sonification Doomed to Fail?" in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*. Newcastle upon Tyne: Department of Computer and Information Sciences, Northumbria University, Jun. 2019, pp. 327–330.

[16] A. Supper, "Sound Information: Sonification in the Age of Complex Data and Digital Audio," *Information & Culture*, vol. 50, no. 4, pp. 441–464, 2015.

[17] S. Barrass and P. Vickers, "Sonification design and aesthetics," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Verlag Berlin, 2011.

[18] A. De Angeli, A. Sutcliffe, and J. Hartmann, "Interaction, usability and aesthetics: what influences users' preferences?" in *Proceedings of the 6th conference on Designing Interactive systems*, 01 2006.

[19] K. Vogt, V. Goudarzi, and R. Parncutt, "Empirical aesthetic evaluation of sonifications," in *Proceedings of the 19th International Conference on Auditory Display (ICAD 2013)*, 2013.

[20] S. D. H. Cornejo, "Towards Ecological, Embodied and User-Centric Design in Auditory Display," in *Proceedings of the 24th International Conference on Auditory Display - ICAD 2018*. Houghton, Michigan: The International Community for Auditory Display, Jun. 2018, pp. 191–196.

[21] G. Leplaitre and I. McGregor, "How to tackle auditory interface aesthetics? discussion and case study," in *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004*, 2004.

[22] P.-J. Maes, J. Buhmann, and M. Leman, "3mo: A Model for Music-Based Biofeedback," *Frontiers in Neuroscience*, vol. 1, 12 2016.

[23] I. Bergstrom, S. Seinfeld, J. Arroyo Palacios, M. Slater, and M. Sanchez-Vives, "Using music as a signal for biofeedback." *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 93, 04 2013.

[24] P. Kantan, E. G. Spaich, and S. Dahl, "A technical framework for musical biofeedback in stroke rehabilitation," *IEEE Transactions on Human-Machine Systems*, 2022.

[25] J. E. Ollen, "A criterion-related validity test of selected indicators of musical sophistication using expert ratings," Ph.D. dissertation, The Ohio State University, 2006.

[26] I.-H. Hsieh and K. Saberi, "Temporal integration in absolute identification of musical pitch," *Hearing Research*, vol. 233, no. 1-2, pp. 108–116, 2007.

[27] S. Barrass, N. Schaffert, and T. Barrass, "Probing preferences between six designs of interactive sonifications for recreational sports, health and fitness," *Proceedings of ISon*, pp. 23–29, 2010.