RESEARCH

Simultaneous but independent spatial associations for pitch and loudness

Sarah Koch[1] · Torsten Schubert[1] · Sven Blankenberger[1]

Abstract

For the auditory dimensions loudness and pitch a vertical SARC effect (Spatial Association of Response Codes) exists: When responding to loud (high) tones, participants are faster with top-sided responses compared to bottom-sided responses and vice versa for soft (low) tones. These effects are typically explained by two different spatial representations for both dimensions with pitch being represented on a helix structure and loudness being represented as spatially associated magnitude. Prior studies show incoherent results with regard to the question whether two SARC effects can occur at the same time as well as whether SARC effects interact with each other. Therefore, this study aimed to investigate the interrelation between the SARC effect for pitch and the SARC effect for loudness in a timbre discrimination task. Participants ($N = 36$) heard one tone per trial and had to decide whether the presented tone was a violin tone or an organ tone by pressing a top-sided or bottom-sided response key. Loudness and pitch were varied orthogonally. We tested the occurrence of SARC effects for pitch and loudness as well as their potential interaction by conducting a multiple linear regression with difference of reaction time (dRT) as dependent variable, and loudness and pitch as predictors. Frequentist and Bayesian analyses revealed that the regression coefficients of pitch and loudness were smaller than zero indicating the simultaneous occurrence of a SARC effects for both dimensions. In contrast, the interaction coefficient was not different from zero indicating an additive effect of both predictors.

Introduction

Several stimulus dimensions show a horizontal, spatial association (Macnamara et al., 2018). Probably one of the most prominent examples are numbers: Small numbers are assumed to be represented left while large numbers are assumed to be represented right on a spatial mental number line (Dehaene et al., 1993; Feigenson et al., 2004; Restle, 1970). Empirical evidence for this assumption stems from the SNARC effect (Spatial-Numerical Association of Response Codes) first investigated by Dehaene et al. (1993). In that study, participants decided whether a presented number was odd or even by pressing a left-sided or right-sided response key. Participants responded faster to small numbers with the left-sided response key compared to responding with a right-sided response key and vice versa for large numbers. In recent decades, a comparable effect has been found for other stimulus dimensions which lead to the general term

SARC effect (Spatial Association of Response Codes, Macnamara et al., 2018).

Another term widely used is the abbreviation SQUARC effect (Spatial-Quantity Association of Response Codes) introduced by Walsh (2003) in the proposal of A Theory of Magnitude (ATOM; Bueti & Walsh, 2009). According to ATOM, the three domains of time, space, and quantity are represented on a common cortical metric, which can also be interpreted as a generalized magnitude representation system (Bonn & Cantlon, 2012). An important prediction of ATOM is the existence of spatial associations for each magnitude dimension. This association is reflected in the SQUARC effect, that is, shorter reaction times to small quantities with a left-sided response than with a right-sided response and vice versa for large quantities (Walsh, 2003, 2015).

In addition, SARC effects have also been shown for the auditory dimensions loudness and pitch. In general, participants respond faster to soft or low tones with a left-sided response key compared to a right-sided response key and vice versa for loud or high tones (Fairhurst & Deroy, 2017; Guilbert, 2020; Hartmann & Mast, 2017; Lega et al., 2020; Lidji et al., 2007; Rusconi et al., 2006). The SARC effects for loudness and pitch also occur in the vertical dimension

✉ Sarah Koch
  sarah.koch@psych.uni-halle.de

1 Department of Psychology, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

indicating that soft or low tones are associated with the spatial information 'bottom' while loud or high tones are associated with the spatial information 'top' (Bruzzi et al., 2017; Fernandez-Prieto et al., 2017; Lega et al., 2020; Lidji et al., 2007; Pitteri et al., 2017; Rusconi et al., 2006). This is in line with the observation, that spatial associations for different stimulus dimensions can occur in several spatial axes such as vertical or radial (i.e. near-far) axes (see Winter et al., 2015 for a review).

SARC effects for pitch and loudness are typically explained by assuming a spatial representation of the corresponding auditory dimension. However, the structure of the assumed spatial representation differs between pitch and loudness. Pitch is assumed to be represented on a two-dimensional spatial helix structure (Shepard, 1982; Ueda & Ohgushi, 1987). Contrary, the SARC effect for loudness is explained by loudness being represented as a magnitude dimension with a linear spatial association (e.g. Bruzzi et al., 2017). This indicates that the effects rely on similar mechanisms but separated representations with different spatial structures. However, little is known how these assumed separated representations relate to each other, that is whether the SARC effects for pitch and loudness can occur simultaneously and independently from each other. Therefore, this study aimed to investigate the interrelation between the SARC effects for pitch and for loudness. For this, we tested whether both SARC effects interact with each other, which would be reflected in a larger SARC effect for pitch for loud tones compared to soft tones and a larger SARC effect for loudness for high tones compared to low tones. For a better distinction between both effects, we will refer to the effects as SPARC effect (Spatial-Pitch Association of Response Codes, Lidji et al., 2007) and as SLARC effect (Spatial-Loudness Association of Response Codes).[1]

Before outlining the assumed interrelation in more detail, we will first describe the main findings for the SPARC and SLARC effects, separately. The SPARC effect depends on the interplay of various factors, namely, musical experience of the participants, the spatial arrangement of response keys, and whether pitch is the task relevant dimension or not. When participants have to classify the pitch of a presented tone relative to a standard pitch, a SPARC effect occurs regardless of the spatial arrangement of the response keys and musical experience (Guilbert, 2020; Lega et al., 2020; Lidji et al., 2007; Rusconi et al., 2006). In contrast, when pitch is not relevant for the task and participants have to respond to another attribute of the tone, for example its timbre, non-musicians show only a SPARC effect with

vertically but not with horizontally aligned response keys. Contrary, musicians still show a SPARC effect in a timbre discrimination task in the horizontal dimension as well as in the vertical dimension (Lega et al., 2020; Lidji et al., 2007; Rusconi et al., 2006).

The occurrence of the vertical SPARC effect is typically explained by the assumption of a mental spatial representation of pitch on a bottom-to-top helix structure (Shepard, 1982; Ueda & Ohgushi, 1987). Low pitches are assumed to be represented bottom while high pitches are assumed to be represented top. Additionally, the assumed representation model takes into account that the relationship between the physical dimension frequency and the perception of pitch is non-linear: An increasing frequency does not only lead to the impression of an increasing pitch height but does also change the perceived quality of the tone. This is referred to as pitch chroma and is indicated by the circular organization of tones within a helix plane. The occurrence of the vertical SPARC effect even under conditions in which pitch is not relevant for the task further indicates that the spatial information is automatically co-activated comparable to the automatic activation in case of the SNARC effect in a parity judgment task (Dehaene et al., 1993), which is in line with the assumption of an innate spatial representation of pitch. Note that an alternative explanation could be the semantic overlap between the response codes and the stimulus codes, as pitch is described in spatial terms in some languages. However, the SPARC effect also occurs in participants whose native language does not describe pitch via spatial terms (Fernandez-Prieto et al., 2017), thus invalidating the semantic overlap explanation.

Contrary to the SPARC effect, the SLARC effect does not dependent on the arrangement of response keys or the relevance of loudness for the task. Although earlier studies did not find a SLARC effect in the horizontal dimension (Ren et al., 2011), the effect was later found by other studies (Chang & Cho, 2015; Fairhurst & Deroy, 2017; Hartmann & Mast, 2017). Furthermore, several studies found the SLARC effect in the vertical dimension (Bruzzi et al., 2017; Fernandez-Prieto et al., 2017). The SLARC effect is not only present when participants have to judge the loudness of a tone relative to a standard tone (Bruzzi et al., 2017; Hartmann & Mast, 2017) but does also occur in timbre discrimination tasks with horizontally (Chang & Cho, 2015) and vertically arranged response keys (Koch et al., 2023). This indicates an automatic activation of the spatial information of loudness comparable to the automatic activation of the spatial information in the case of pitch.

Typically, the SLARC effect is explained by assuming that loudness is represented as a magnitude according to ATOM (Bueti & Walsh, 2009; Walsh, 2003), as suggested for instance by Bruzzi et al. (2017). The SLARC effect found in previous studies is in line with this prediction as loud

---

[1] Please note, that the term SLARC effect is also used in the context of spatial associations for linguistic material (Abbondanza et al., 2021).

tones, that is tones with high intensity, are associated with right and soft tones associated with left (Hartmann & Mast, 2017). Furthermore, ATOM predicts interactions between magnitude dimensions (Walsh, 2003, 2015) and several studies found that loudness interacts with other magnitude dimensions like numerical magnitude (Alards-Tomalin et al., 2015; Hartmann & Mast, 2017; Heinemann et al., 2013) or physical size (Smith & Sera, 1992; Sutherland et al., 2014; Takeshima & Gyoba, 2013) which also supports the assumption that loudness is represented as a magnitude. Furthermore, loudness is a prothetic or quantitative dimension (Stevens, 1957; Stevens & Galanter, 1957), which is an important theoretical prerequisite for a dimension to be considered part of ATOM (Walsh, 2003, 2015). Taken together, assuming that loudness is represented as a magnitude in the sense of ATOM, the SLARC effect could be considered as an instance of the general SQUARC effect. This assumption is also supported by the notion that the SLARC effect seems to be continuous (Koch et al., 2023) which contradicts the most prominent alternative explanation, namely the polarity correspondence principle (Chang & Cho, 2015; Proctor & Cho, 2006).

So far, empirical evidence suggests that the SPARC and SLARC effects are due to two different spatial representations but a direct empirical test of this assumption is still missing. In addition, previous studies investigating spatial associations for spoken number words have found that the SPARC and SLARC effects interact with the already mentioned SNARC effect in a way that contradicts several theoretical assumptions. Numbers are assumed to be represented as a magnitude in terms of ATOM (Bueti & Walsh, 2009; Walsh, 2003, 2015) and therefore would share a magnitude representation with loudness. Pitch, on the other hand, is explicitly excluded from the conceptualization of ATOM because it is a metathetic or qualitative dimension (Stevens, 1957; Stevens & Galanter, 1957). Based on these assumptions, two interaction patterns between the SLARC, SPARC, and SNARC effects are plausible. First, from the premise that an interaction between spatial associations indicates a common origin, one would expect that the SLARC and SNARC effects should interact. This interaction could be reflected in a larger SLARC effect for large spoken numbers compared to small numbers as well as a larger SNARC effect for loud spoken number words compared to soft spoken number words. The SPARC and SNARC effects should be independent of each other. Alternatively, since a shared representation does not rule out purely additive effects (Sternberg, 1969), a second possible scenario could be that loudness and numbers share a common representation, but that the SLARC and SNARC effect simply do not interact. Importantly, the SPARC and SNARC effects should still not interact. Results from previous studies contradict both scenarios: While the SLARC effect does not interact with the

SNARC effect (Hartmann & Mast, 2017), the SPARC effect does interact with the SNARC effect (Fischer et al., 2013; Weis et al., 2015, 2016).

In a study by Hartmann and Mast (2017), participants heard spoken number words and had to classify the numerical value, loudness level, or parity. There was no interaction between the SPARC effect and the SLARC effect. Additionally, both effects were limited to the dimension-related task. The SNARC effect only occurred in the parity and number judgment task while the SLARC effect was limited to the loudness judgment task. However, both effects are known to occur even when number magnitude or loudness are irrelevant (Chang & Cho, 2015; Fias, 2001; Koch et al., 2023; for a review for the SNARC effect see Wood et al., 2008). This raises the question of whether two SARC effects can generally occur simultaneously.

Indeed, studies that found an interaction between the SPARC effect and the SNARC effect also found a simultaneous occurrence of both effects (Weis et al., 2015, 2016). However, the results are not entirely consistent, which might be due to different experimental setups. For example, Fischer et al. (2013) investigated the SPARC effect and the SNARC effect in a pitch discrimination task and a number discrimination task with diagonally arranged response keys. Both effects only occurred when the corresponding dimension was task-relevant. Crucially, there was a reversed SNARC effect for the SPARC incompatible trials but not for the SPARC compatible trials. In the studies conducted by Weis and colleagues (Weis et al., 2015, 2016), participants had to classify either the numerical value, pitch, or parity. Participants responded faster in SNARC compatible and SPARC compatible trials compared to incompatible trials regardless of the task. Furthermore, there was a significant interaction between SPARC compatibility and SNARC compatibility. The authors concluded that both effects share a common automatic decision mechanism and further suggested that this mechanism might be based on a common representation of pitch and numbers in the sense of ATOM (Weis et al., 2016). However, as already mentioned, pitch is explicitly excluded from the conceptualization of ATOM due to its metathetic or qualitative characteristic (Walsh, 2003, 2015). From a theoretical point of view, the interaction between SPARC and SNARC compatibility cannot be explained in terms of a common representation in the sense of ATOM.

Taken together, previous findings contradict predictions regarding potential interactions between SARC effects. In addition, it is unclear under which circumstances and for which dimensions two SARC effects can occur simultaneously. Therefore, the aim of our study was twofold. First, we wanted to investigate whether the SLARC effect occurs simultaneously with another spatial association, namely the SPARC effect. We used a timbre discrimination task in which participants had to decide whether a single tone

was a violin tone or an organ tone while pitch and loudness were varied orthogonally. Neither loudness nor pitch were relevant for the task, which allowed us to investigate whether both effects would show an automatic, simultaneous occurrence. Furthermore, timbre is equally strong related to both task-irrelevant dimensions, which is not the case for parity as used in previous studies (Hartmann & Mast, 2017; Weis et al., 2015, 2016), which is stronger related to the numerical value than to pitch or loudness. As a second aim, we wanted to test whether the interaction between the SPARC and SNARC effects (Weis et al., 2015, 2016) generalizes to the interrelation between the SLARC and the SPARC effects. The interaction between the SPARC and SNARC effects is explained by a shared representation according to ATOM (Weis et al., 2016). If this is the case, then the interaction should generalize to other magnitude dimensions as well. The SLARC effect is explained by an assumed magnitude representation of loudness (Bruzzi et al., 2017), and therefore one would also expect an interaction between the SLARC effect and the SPARC effect. This would be reflected in a larger SPARC effect for loud tones compared to soft tones as well as a larger SLARC effect for high tones compared to low tones.

To test the simultaneous occurrence as well as a possible interaction between both effects, we conducted a multiple linear regression, with the difference of reaction time between top-sided and bottom-sided responses (dRT) as dependent variable and loudness and pitch as predictors (Fias et al., 1996). We predicted that the SPARC effect and the SLARC effect would occur simultaneously indicated by a negative regression coefficient for loudness as well as for pitch. Previous studies found that SARC effects for two prothetic dimensions did not occur simultaneously (Hartmann & Mast, 2017; Vellan & Leth-Steensen, 2022; Weis et al., 2018) whereas SARC effects for one prothetic and one metathetic dimension occured simultaneously (Weis et al., 2015, 2016). Because pitch and loudness are regarded as metathetic and prothetic dimensions, respectively (Stevens & Galanter, 1957), SARC effects for both dimensions should occur at the same time. With regard to our second research aim, we predicted that if an interaction between both effects occurred, it should be reflected in larger dRT differences between soft and loud tones which are high in pitch compared to tones which are low in pitch.

Methods

The study's design as well as our hypothesis and analysis plan were preregistered on aspredicted.org (https://aspredicted.org/CG8_V9Q).

Participants

$N = 38$ healthy students from Martin Luther University Halle-Wittenberg participated for course credit. Age was restricted to a range from 18 to 35 years to ensure that age-related hearing impairments were negligible between participants. Professional musicians (i.e. people who study or have studied music or work full-time or part-time as musician) were not allowed to participate in the study. The planned sample size was $N = 36$ based on a power analysis for a one-tailed one-sample t-test with a significance level of $\alpha = 0.05$, Power $1 - \beta = 0.9$ and a medium effect size of $d_z = -0.5$. After we reached our determined sample size, we checked the mean error rate and excluded two participants because their error rate exceeded 15%. In line with our preregistration, we collected data from two additional participants. Therefore, data from $N = 36$ participants ($N = 31$ female) were included in the final data analysis. The mean age was $M = 21.8$ years ($SD = 3$). $N = 33$ participants reported being right-handed. All participants reported having normal or corrected-to-normal vision and no participant reported any hearing impairments. Informed written consent was obtained from all individual participants included in the study.

Materials

The auditory material consisted of 50 different tones with five pitch levels, five loudness levels, and two timbres. Pitch was operationalized by frequency in Hertz and pitch levels ranged from from 261 Hz (F#3) to 523 Hz (C5) with 3 half-tone steps between each pitch level. Loudness levels ranged from 45 to 65 phon with 5 phon between each level. All pitch-loudness combinations were realized as violin tone and as organ tone. Auditory stimuli were first synthesized with the software LMMS (Version, 1.2.0, Junghans & Giblock, 2019). Pitch levels of the tones were then analyzed via spectral analyses in Audacity (Version 3.2.0, The Audacity Team, 2019) and adjusted to the aimed fundamental frequency if necessary. Because the subjective loudness perception depends on sound pressure level and frequency (Fletcher & Munson, 1933), we controlled the loudness level for each frequency as followed. First, we adjusted the sound pressure level for each pitch level according to the isophone curves from the ISO norm for pure tones (International Organization for Standardization, 2003). Afterwards, tones were presented to $N = 4$ independent participants, who adjusted the amplitude of the tones according to their subjective loudness impression, so that all tones from one loudness level sounded equally loud to them regardless of pitch and timbre. Tone presentation and amplitude adjustment was done directly in Audacity. Participants could repeat the tones and adjust the amplitude of each tone as often as they

wished. Each final sound file had a sample rate of 48 kHz and a duration of 800 ms.

In the experiment, tones were presented via headphones (Sennheiser HD471). Participants responded to the tones by pressing a top-sided or bottom-sided response key, which were vertically aligned on a custom-built response box. Keys were 1 cm and 16 cm above the table surface. Keys were vertically aligned because several studies found the SPARC and SLARC effects occur for non-musicians in a timbre discrimination task with vertically aligned response keys (Koch et al., 2023; Lega et al., 2020; Lidji et al., 2007; Rusconi et al., 2006). Stimulus presentation and data recording was realized via the software PsychoPy (Version 2021.2.3, Peirce et al., 2019).

Procedure

Each trial started with the presentation of a fixation cross in the middle of the screen for 500 ms. After a foreperiod of 500 ms, the tone was presented for a maximum of 800 ms or until participants made a response. Because there was no time window, participants could press the response key after the tone ended. After participants gave their response, an inter-trial interval (ITI) of a uniformly distributed duration between 1 and 2 s was presented before the next trial started. Participants were instructed to classify the presented tone as violin tone or organ tone as fast and accurate as possible by pressing the top-sided or bottom-sided response key. They were not told that the tones had different loudness and pitch levels but to make them familiar with the auditory material, participants heard the combinations of the loudest, softest, highest, and lowest tones in both timbres during the instruction phase.

Response mapping was varied within participants and between four sessions. Order of response mapping was counterbalanced across participants. Half of the participants responded in the first and second session with the top-sided response key to the violin tone and with the bottom-sided response key to the organ tone and vice versa in the third and fourth session. The other half of participants had the opposite order of response mapping. Key-hand mapping was also varied between participants. Half of the participants pressed the top-sided response key with their right thumb and the bottom-sided response key with their left thumb, the other half had the opposite mapping. Participants were randomly assigned to one of the four condition groups with $N = 9$ participants in each group.

In each session, all 50 tones were presented 12 times resulting in 600 experimental trials divided into 12 blocks à 50 experimental trials plus three warm-up trials in the beginning of each block. Participants performed 50 training trials in the beginning of each session in which they received feedback about their reaction time and correctness of the response. Each session took between 45 and 50 min.

At the end of the fourth session, participants completed two subscales from the German version of the Goldsmith's Musical-Sophistication Index (Gold-MSI, Müllensiefen et al., 2014; Schaal et al., 2014). Although professional musicians were not allowed to participate in the study, we assumed that some participants might be familiar with playing an instrument. Some studies rely on cut-off criteria to define non-musicality, for example not playing an instrument for a certain amount of years (e.g. Lidji et al., 2007; Weis et al., 2016). However, these cut-off criteria are often not theoretical or empirical justified. Therefore, we assessed musical experience with a standardized questionnaire to investigate potential influences on the SPARC effect or SLARC effect. For our study, we used the subscales "perception" (9 items) and "musical training" (7 items) with a Cronbach's α of $\alpha = 0.8$ for the perception scale and $\alpha = 0.91$ for the musical training scale ($\alpha = 0.83$ and $\alpha = 0.88$ in the study of Schaal et al. (2014), respectively).

Data analysis

We used R (Version 4.1.2; R Core Team, 2021) and the R-packages *afex* (Version 1.1.1; Singmann et al., 2022), *BayesFactor* (Version 0.9.12.4.4; Morey & Rouder, 2022), *papaja* (Version 0.1.1; Aust & Barth, 2022), and *tidyverse* (Version 1.3.1; Wickham et al., 2019) for all our analyses. Reaction times from incorrect trials (6.9%) were discarded from further analyses. We calculated the trimmed mean reaction times with a trimming amount of 20% (Rosenberger & Gasko, 1983) and the mean error rate for each participant and within-subject condition. In addition, we calculated the mean difference of reaction times ($dRT = RT_{top} - RT_{bottom}$) for each participant, loudness, frequency, and timbre (Fias et al., 1996). For the ANOVA results, *p*-values corrected according to Geisser-Greenhouse (GG) will be reported in case of violations of the sphericity assumption. For the dRT analyses, we will report results from frequentist and Bayesian analyses.
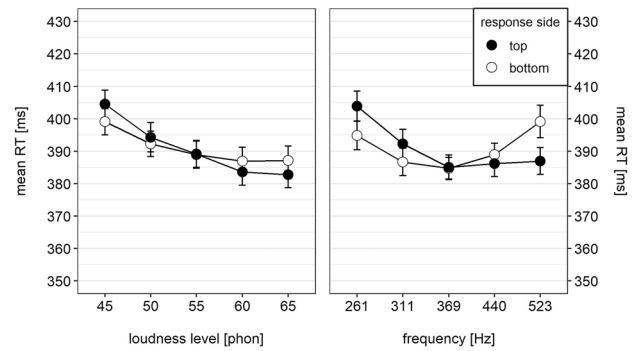
Results

ANOVA results

In a first step, we conducted a mixed 5 (loudness) × 5 (frequency) × 2 (timbre) × 2 (response side) × 2 (order of response side) × 2 (key-hand mapping) ANOVA with order of response mapping and key-hand mapping as between-subjects factors and the remaining factors as within-subjects factors. Overall, mean reaction time decreased with increasing loudness with 402 ms, 393 ms, 389 ms, 385 ms,
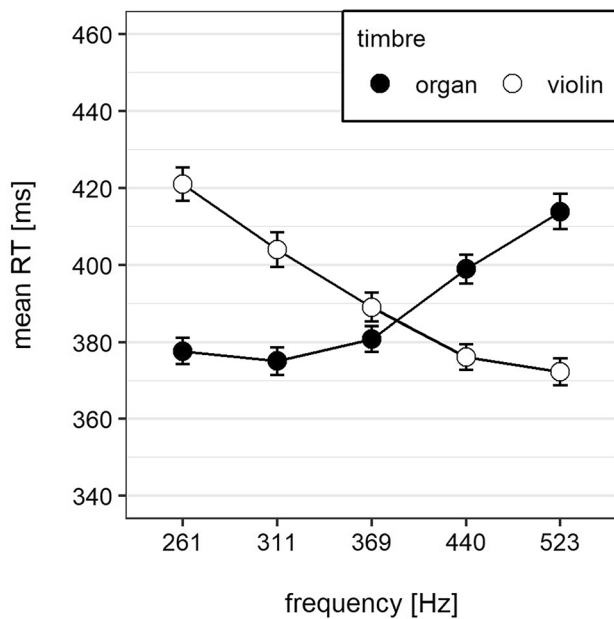
Fig. 1 Mean reaction time as a function of frequency and timbre. Error bars represent 95% within-subjects CI (Morey, 2008)


Fig. 2 Mean reaction time as a function of response side and loudness (left panel) and response side and pitch (right panel). Error bars represent 95% within-subjects CI (Morey, 2008)

and 384 ms from softest to loudest loudness level. This main effect of loudness was significant, $F(4, 128) = 44.14$, $MSE = 1211.94$, $p < 0.001$ (GG), $\hat{\eta}_G^2 = 0.007$. There was also a significant main effect of frequency, $F(4, 128) = 22.16$, $MSE = 1807.04$, $p < 0.001$ (GG), $\hat{\eta}_G^2 = 0.004$: Participants responded fastest to tones with a frequency of of 369 Hz (385 ms) and their reaction time increased with decreasing (389–399 ms) and increasing frequency (387–393 ms). Additionally, there was also a significant interaction between frequency and timbre, $F(4, 128) = 81.16$, $MSE = 7074.73$, $p < 0.001$ (GG), $\hat{\eta}_G^2 = 0.040$. Participants responded faster to high violin and low organ tones compared to low violin tones and high organ tones (see Fig. 1).

With regard to our hypothesis, there was a significant interaction between loudness and response side, $F(4, 128) = 5.07$, $MSE = 697.48$, $p = 0.002$ (GG), $\hat{\eta}_G^2 = 0.001$ as well as a significant interaction between frequency and response side, $F(4, 128) = 12.07$, $MSE = 1988.93$, $p < 0.001$ (GG), $\hat{\eta}_G^2 = 0.002$. Both interactions are depicted in Fig. 2. Participants responded faster to soft or low tones with the bottom-sided response key compared to the top-sided response key. The opposite holds true for loud and high tones, respectively. For both dimensions, there was also a significant three-way interaction with timbre, with $F(4, 128) = 3.04$, $MSE = 848.39$, $p = 0.037$ (GG), $\hat{\eta}_G^2 = 0.000$ for the loudness × response side × timbre interaction and $F(4, 128) = 3.49$, $MSE = 670.29$, $p = 0.017$ (GG), $\hat{\eta}_G^2 = 0.000$ for the frequency × response side × timbre interaction. A more detailed analysis of both effects is carried out in the dRT analyses in the next section.

Response side interacted with both between-subject factors, order of response mapping, $F(1, 32) = 5.28$, $MSE = 3, 309.02$, $p = 0.028$, $\hat{\eta}_G^2 = 0.001$, and key-hand-assignment, $F(1, 32) = 6.14$, $MSE = 3309.02$, $p = 0.019$, $\hat{\eta}_G^2 = 0.001$. Participants who responded in the first two sessions to the violin tone with the top-sided response key were in general faster when responding with the top-sided response key (393 ms vs. 396 ms). Contrary, participants who started with the opposite mapping were in general faster when they pressed the bottom-sided response key compared to a top-sided response key (385 ms vs. 389 ms). Additionally, participants who pressed the bottom-sided response key with their right thumb were faster when pressing the bottom-sided response key compared to the top-sided response key (388 ms vs. 393 ms) and vice versa for participants who had the opposite key-hand mapping (387 ms vs. 392 ms). Because 33 out of 36 participants reported being right-handed, this interaction probably reflects the effect of the dominant hand on reaction times. There was also a significant interaction between timbre, response side, and order of response mapping, $F(1, 32) = 31.22$, $MSE = 24, 385.61$, $p < 0.001$, $\hat{\eta}_G^2 = 0.034$. Participants were in general faster in the response mapping of the third and fourth sessions compared to the response mapping of the first and second session, which lead to the significant three-way interaction. Lastly, the three-way interaction between response side, frequency, and order of response mapping was significant, $F(4, 128) = 4.21$, $MSE = 1988.93$, $p = 0.019$ (GG), $\hat{\eta}_G^2 = 0.001$. All other main effects and interactions were non-significant, $p > 0.05$.

Results from dRT analyses

Mean dRT as a function of loudness and frequency averaged across the other predictor is illustrated in Fig. 3. For both dimensions, mean dRT decreased with an increasing value of the predictor. This holds true when visualizing the
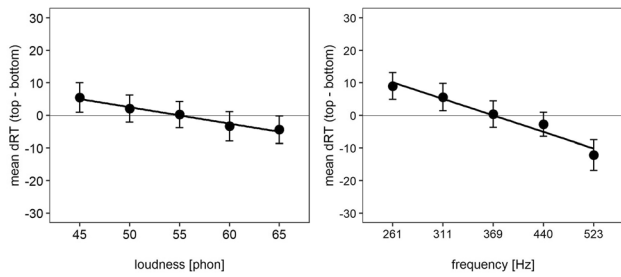
Fig. 3 Mean dRT and regression line as a function of loudness (left panel) and frequency (right panel) averaged over the other dimension. Error bars represent 95% within-subjects CI (Morey, 2008)
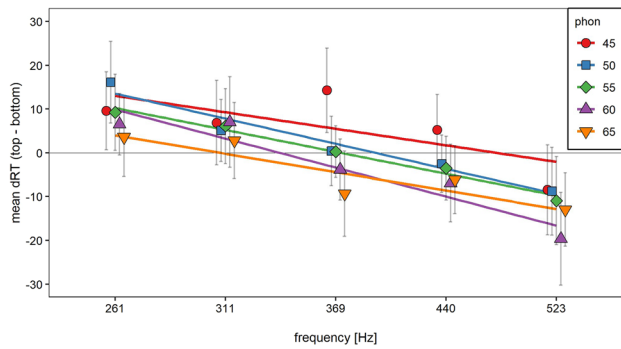
Fig. 4 Mean dRT and regression lines as a function of loudness and frequency. Error bars represent 95% within-subjects CI (Morey, 2008)

relationship between dRT and the loudness × frequency interaction (see Fig. 4). Mean dRT decreased with increasing loudness level for all frequencies and mean dRT decreased with increasing frequency for all loudness levels. From a visual inspection, there was no clear interaction trend visible.

To test the main and interaction effects, we conducted a multiple linear regression with loudness and frequency as predictors and dRT as dependent variable. We allowed the predictors to interact, although the non-significant interaction between response side, loudness, and frequency already indicated an additive effect. Because of

the different scales for loudness and frequency, we first standardized the predictors and the dependent variable to facilitate comparisons between the regression coefficients. Regression coefficients were estimated separately for each participant and timbre and then tested against zero via one-sample t-tests (Lorch & Myers, 1990). We used one-tailed (preregistered), one-sample t-tests for the main effects and two-tailed (preregistered), one-sample t-test for the inter-action coefficients. Because timbre significantly interacted with several factors in the previously conducted ANOVA, we followed our preregistration protocol and tested the overall regression coefficients as well as regression coef-ficients separated for both timbres. This procedure lead to nine one-sample t-tests in total. To avoid multiple com-parison problems, we conducted a sequential Bonferroni correction, namely Rom's procedure to adjust significance levels (Olejnik et al., 1997). Results of the t-tests as well as the adjusted significance levels are depicted in Table 1. Regression coefficients of the interaction term were not significantly different from zero in all three conditions. The regression coefficients for frequency were signifi-cantly smaller than zero in all three conditions. Results for the regression coefficients for loudness were mixed. While in the overall and organ condition, regression coef-ficients were significantly smaller than zero, this was not the case in the violin condition.

Results from the frequentist analysis indicate that there was no significant interaction between loudness and frequency. To further support the interpretation of an absent effect, we also conducted Bayesian one-sam-ple t-tests with a Cauchy prior distribution with a prior width of $r = 1/\sqrt{2}$ (Rouder et al., 2009). To resemble the preregistered one-tailed and two-tailed testing, we used a truncated distribution for testing the regression coeffi-cients of loudness and frequency, and the full distribu-tion for testing the interaction coefficients (for a detailed description of this approach, see Morey & Rouder, 2011). Due to multiple testing, we corrected the prior probability of the H_0 according to Westfall et al. (1997)

Table 1 Results from nine one-sample t-tests and corrected significance levels for the overall condition as well as separated for each timbre

Condition	Predictor	M	95% CI	t (35)	p	$\alpha_{corrected}$
Violin	Interaction	0.01	$[-0.04, 0.05]$	0.30	0.768	0.05
Overall	Interaction	-0.01	$[-0.06, 0.04]$	-0.32	0.749	0.02
Organ	Interaction	-0.02	$[-0.05, 0.02]$	-0.99	0.330	0.02
Violin	Loudness	-0.06	$[-\infty, -0.01]$	-2.10	0.021	0.01
Organ	Loudness	-0.07	$[-\infty, -0.03]$	-3.03	0.002	0.01
Overall	Loudness	-0.12	$[-\infty, -0.06]$	-3.19	0.001	0.01
Organ	Frequency	-0.12	$[-\infty, -0.06]$	-3.46	0.001	0.01
Overall	Frequency	-0.25	$[-\infty, -0.14]$	-3.90	<0.001	0.01
Violin	Frequency	-0.14	$[-\infty, -0.08]$	-3.92	<0.001	0.00

Table 2 Results from the Bayesian one-sample *t*-tests for the overall condition as well as separated for each timbre

Condition	Predictor	M	95% HDI	BF_{10}/BF_{-0}	Prior odds$_{corr}$	Posterior odds$_{corr}$
Violin	Interaction	0.01	[−0.04, 0.05]	0.19	0.26	0.05
Overall	Interaction	−0.01	[−0.06, 0.04]	0.19	0.26	0.05
Organ	Interaction	−0.01	[−0.04, 0.02]	0.28	0.26	0.07
Violin	Loudness	−0.05	[−0.10, 0.00]	2.47	0.1	0.25
Organ	Loudness	−0.07	[−0.11, −0.02]	16.6	0.1	1.67
Overall	Loudness	−0.11	[−0.19, −0.04]	24.2	0.1	2.43
Organ	Frequency	−0.12	[−0.19, −0.05]	45.2	0.1	4.55
Overall	Frequency	−0.23	[−0.36, −0.11]	138	0.1	13.88
Violin	Frequency	−0.13	[−0.20, −0.06]	146	0.1	14.70

HDI = Highest Density Interval. Estimation errors of the Bayes factors were all smaller than 5% and are therefore not reported. Prior odds and posterior odds are based on the corrected prior probability for H_{0i}

leading to $p(H_{0i}) = 0.5^{1/3} = 0.79$ for two-sided tests and $p(H_{0i}) = 1/3^{1/6} = 0.83$ for one-sided tests.[2]

Mean regression coefficients as well as the corresponding Bayes factors (BF_{10} for two-sided, BF_{-0} for one-sided tests) and posterior odds are depicted in Table 2. Note that *M* reports a point estimate based on the resulting posterior distribution and can therefore differ from the empirical means. For all interaction coefficients, BF_{10} was smaller than 0.3 indicating that the data was more likely under the H_0, that is, that the regression coefficient is not different from zero. For the main effects, all BF_{-0} but one were larger than 10 indicating that the data was more likely under H_1, that is, that the regression coefficient is smaller than zero. For the regression coefficient for loudness in the violin condition, BF_{-0} was smaller than 3 and therefore indicated that data was 2.47 times more likely under the H_1. Additional robustness checks further showed that the Bayes Factors remained comparable for a range of different prior widths (see "Appendix" for the full robustness check).

To investigate whether musical training or perceptual abilities had an influence on the SPARC effect or the SLARC effect, we calculated Pearson's correlation coefficients between the mean score for each scale and the overall regression coefficients. Mean scores were $M_{perception} = 5.11$ ($SD = 0.81$) and $M_{training} = 3.20$ ($SD = 1.58$). There was a significant correlation between mean scores for both scales, $r = 0.70$, 95% CI [0.49, 0.84], $t(34) = 5.76$, $p < 0.001$. Neither the mean score of perception nor the mean score of musical training did significantly correlate with any of the regression coefficients ($ps > 0.05$). Therefore, we resigned from further analysis.

Discussion

The first aim of this study was to investigate whether the SPARC effect and the SLARC effect occur simultaneously in a timbre discrimination task, that is, when loudness and pitch are irrelevant for the task. Indeed, loudness as well as pitch interacted with response side: Participants responded faster to high and loud tones when responding with the top-sided response key compared to the bottom-sided response key and vice versa for soft and low tones. The dRT analyses further revealed, that mean dRT linearly decreased with increasing loudness as well as with increasing pitch. These results show that both the SPARC effect and the SLARC effect occurred, supporting our first hypothesis regarding the simultaneous occurrence of both effects. A second aim of this study was the investigation of a potential interrelation between the SPARC effect and the SLARC effect indicated by an interaction between both effects. Contrary to our second hypothesis, the predictors loudness and pitch did not interact in the dRT analyses and the effects were purely additive.

Previous studies investigated either the SLARC effect or the SPARC effect in a timbre discrimination task (Koch et al., 2023; Lega et al., 2020; Lidji et al., 2007; Rusconi et al., 2006). The results from our study did not only replicate these effects, but also showed that both effects can occur simultaneously. Our SPARC effect was numerically smaller compared to results from other studies. This can easily be explained by the use of a limited pitch range in our experiment compared to the pitch ranges used in previous studies (e.g. Lidji et al., 2007; Rusconi et al., 2006). As loudness and pitch were both task-irrelevant, the results indicate a simultaneous and automatic activation of the spatial information in both dimensions. Additionally, the continuous linear decrease of dRT with increasing loudness level indicates a continuous spatial representation rather than a categorization as it would be predicted by, for example, the polarity correspondence principle (Proctor & Cho, 2006).

---

[2] Please note, that the approach for one-sided leads to two rival hypotheses (van Doorn et al., 2020).

The occurrence of SARC effects even when the corresponding dimension is not relevant for the task, is generally considered an indication of automatic activation of the implicit spatial information for the corresponding dimension (Dehaene et al., 1993; Weis et al., 2015). However, the use of bimanual responses may induce a spatial bias, and the processing of the spatial information would no longer be considered implicit (Shaki & Fischer, 2018; Sixtus et al., 2019). These studies used non-lateralized responses, meaning participants responded with a single response key in a go/no-go task, and therefore processed the magnitude information and spatial information implicitly. In these paradigms, participants did not respond faster when the number magnitude and horizontal spatial information matched. Therefore, it was interpreted that the horizontal SNARC effect may not reflect a spatial representation but rather a spatial processing bias. In contrast to the horizontal spatial information, a reaction time benefit was observed when number magnitude and vertical spatial information matched. This suggests that the vertical association may be inherently linked to the concept of magnitude (Shaki & Fischer, 2018; Sixtus et al., 2019). Regarding pitch and loudness, a next step may be to investigate, whether the vertical associations of these auditory dimensions still occur, when the spatial information is processed implicitly in a setting with non-lateralized responses.

The results show that the SLARC effect can indeed occur simultaneously with another SARC effect, in this case the SPARC effect. The question remains why this was not the case when investigating the SLARC effect and the SNARC effect (Hartmann & Mast, 2017). One possible explanation might be an influence of the task. Fischer et al. (2013) argued, that in the case of two competing dimensions, a potential SARC effect might only arise for the dimension which is relevant for the task. Although loudness and number magnitude are both irrelevant in a parity judgment task as used by Hartmann and Mast (2017), parity is stronger related to the numerical value of a number than to loudness. However, in the present experiment, the task-relevant dimension timbre might be equally strong related to pitch and loudness and therefore a comparable strong automatic activation of the spatial information in both dimensions might have been possible. On the other hand, Weis and colleagues (2015, 2016) also used a parity judgment task and did find simultaneous SARC effects.

For our study, separate frequentist and Bayesian analyses showed that timbre had only a negligible influence on the SPARC and SLARC effect: Both effects occurred in most of the timbre conditions. Nevertheless, the interaction patterns involving timbre partially differed for pitch and loudness.

There was a significant interaction between timbre and pitch with shorter reaction times for high violin and low organ tones compared to low violin and high organ tones. This interaction pattern is comparable to the timbre-pitch interaction found by Melara and Marks (1990). Loudness and timbre did not interact in our study, indicating a slightly different influence of timbre on the processing of pitch and loudness. However, even though timbre might not be completely equally related to pitch and loudness in our study, it did not influence the SPARC and SLARC effect.

Another explanation for discrepancies with regard to the simultaneous occurrences of SARC effect could be that SARC effects for prothetic dimensions do not occur simultaneously in general. Previous studies suggest that this is at least the case for the SNARC effect and the SARC effect for physical size (Vellan & Leth-Steensen, 2022; Weis et al., 2018). However, as these studies used either a number or a size discrimination task, the non-simultaneous occurrence might be due to the different relevance of the dimensions for the task (Fischer et al., 2013). Therefore, and because empirical evidence of concurrent SARC effects is rare, this explanation should be taken with caution. Further research is needed on the simultaneous occurrence of different SARC effects and how this relates to metathetic and prothetic dimensions.

The interaction between the SPARC effect and SNARC effect found by prior studies (Fischer et al., 2013; Weis et al., 2015, 2016) did not generalize to the SLARC effect in our study. This indicates that the interaction between the SPARC effect and the SNARC effect was not due to a shared representation in the sense of ATOM as some authors suggested (Weis et al., 2016). If this would have been the case, the interaction should have generalized to the SLARC effect, as loudness is suggested to be represented as a magnitude in the sense of ATOM (Bruzzi et al., 2017; Hartmann & Mast, 2017). Instead, other mechanisms might have been responsible for the interdependence between the SPARC effect and SNARC effect, for example, sharing a common central processes as already mentioned by Weis et al. (2016).

The lack of an interaction between the SPARC and the SLARC effect supports the assumption that loudness and pitch are represented separately. In addition, the continuous linear decrease of dRT with increasing pitch and loudness indicates that both distinct representations may be continuous. According to Lidji et al. (2007), the SPARC effect may rely on a spatial representation as proposed in former representational models of musical pitch (Shepard, 1982; Ueda & Ohgushi, 1987); while Bruzzi et al. (2017) suggest that the

SLARC effect is due to a generalized magnitude representation of loudness according to ATOM (Walsh, 2003).

Models aiming to describe the mental representation of pitch assume that pitch is represented spatially on a helix structure (Shepard, 1982; Ueda & Ohgushi, 1987). This assumption takes into account that an increase in frequency does not only lead to an increase in perceived pitch height but also to a change of the perceived pitch chroma. Two pitches with the same pitch chroma but from different octaves, for example C4 (261 Hz) and C5 (523 Hz), are considered subjectively more similar than two pitches with different chromas but closer frequencies, such as C4 (261 Hz) and F4 (349 Hz). Nevertheless, while pitch chroma is assumed to be represented circular, the helix structure comprises a constant vertical increase in pitch height. Therefore, even musical tones with the same pitch chroma differ in their pitch height. Thus, a spatial helix representation would still predict a continuous decrease of dRT with increasing frequency, similar to the result pattern in the current study.

A continuous decrease of dRT would be also in line with the assumption of a one-dimensional, linear spatial representation of pitch, comparable to the representation of loudness. However, this representation would not be considered a magnitude representation according to ATOM, because pitch is a metathetic dimension (Stevens, 1957; Stevens & Galanter, 1957), and therefore not part of the generalized magnitude representation system according to ATOM (Walsh, 2003, 2015). Nevertheless, the question whether the SPARC effect relies on a helix structure, or another one-dimensional spatial representation remains, and results from the current study do not allow to distinguish these spatial representation structures. Future studies should address the question whether reaction times indicating a SPARC effect also indicate a spatially organized helix structure of the underlying representation, for example by taking into account the influence of pitch similarity on reaction times in same-different judgments (Cohen Kadosh et al., 2008).

In contrast to pitch, loudness may be represented as a magnitude in the sense of ATOM (Bueti & Walsh, 2009; Walsh, 2003). In this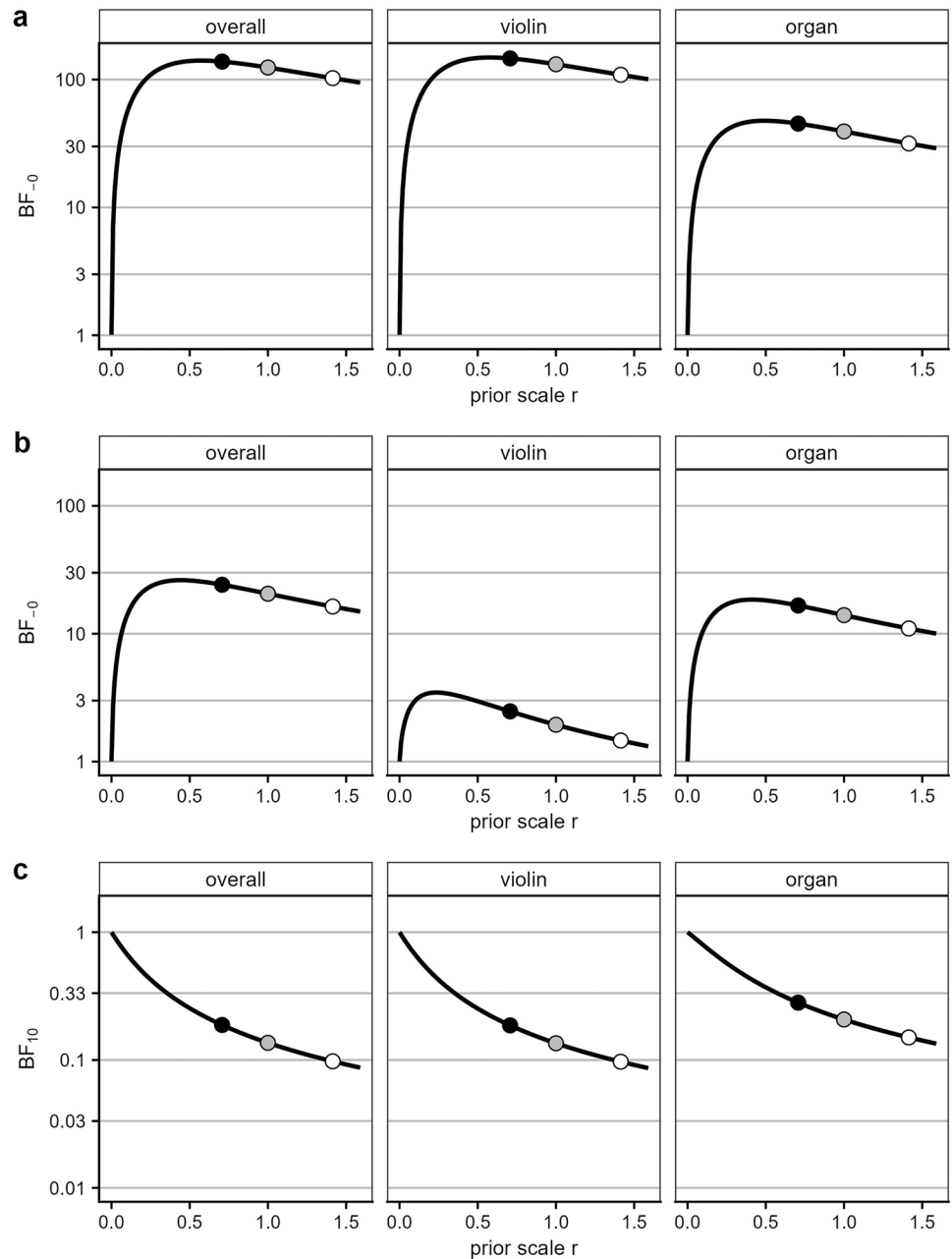 case, the SLARC effect would be an instance of the more general SQUARC effect. This assumption is supported by the prothetic character of loudness (Stevens, 1957; Stevens & Galanter, 1957) and by interactions between loudness and other ATOM-related magnitudes (Alards-Tomalin et al., 2015; Hartmann & Mast, 2017; Heinemann et al., 2013; Takeshima & Gyoba, 2013). The question remains whether a vertical SLARC effect is in line with this interpretation. The direction of the spatial association in the context of ATOM is not narrowed to the horizontal dimension. Furthermore, it is assumed that numbers—one of ATOM's most prominent quantity dimension—are also spatially represented in the vertical dimension (Aleotti et al., 2023; Ito & Hatta, 2004; see Winter et al., 2015 for a review). This vertical spatial association might be present in other magnitudes as well.

In conclusion, our study has shown that the SPARC effect and the SLARC effect occur simultaneously, but appear to be independent of each other. This supports the interpretation that both effects are due to separate spatial representations. Furthermore, our study extended the findings on simultaneous SARC effects and showed that the implicit spatial information of two dimensions can be automatically activated simultaneously. Whether and how this is influenced by task characteristics or by specific characteristics of the dimensions (e. g. the distinction between prothetic and metathetic dimensions) needs to be investigated in further research. In addition, future research could help to understand the complex patterns of interaction between different SARC effects and what leads to interdependencies between spatial associations in different dimensions.

Appendix

See Fig. 5.

Fig. 5 Bayes Factors as a function of prior scale of the Cauchy distribution for the regression coefficients of pitch (**a**), loudness (**b**), and interaction (**c**). *Note*: Points indicate Bayes Factors at prior scale r = $\sqrt{1/2}$ (black), r = 1 (grey), and r = $\sqrt{2}$ (white)

Declarations

ethic committee of the DGPs („Deutsche Gesellschaft für Psychologie").

Informed consent Informed written consent was obtained from all individual participants included in the study.

References

Abbondanza, M., Rinaldi, L., Foppolo, F., & Marelli, M. (2021). The mental representation of nonnumerical quantifiers: The Spatial-Linguistic Association of Response Codes (SLARC) effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(12), 2021–2028. https://doi.org/10.1037/xlm0001037

Alards-Tomalin, D., Walker, A. C., Shaw, J. D. M., & Leboe-McGowan, L. C. (2015). Is 9 louder than 1? Audiovisual cross-modal interactions between number magnitude and judged sound loudness. *Acta Psychologica, 160*, 95–103. https://doi.org/10.1016/j.actpsy.2015.07.004

Aleotti, S., Massaccesi, S., & Priftis, K. (2023). The SNARC effect: A preregistered study on the interaction of horizontal, vertical, and sagittal spatial–numerical associations. *Psychological Research Psychologische Forschung, 87*(4), 1256–1266. https://doi.org/10.1007/s00426-022-01721-8

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

Bonn, C. D., & Cantlon, J. F. (2012). The origins and structure of quantitative concepts. *Cognitive Neuropsychology, 29*(1–2), 149–173. https://doi.org/10.1080/02643294.2012.707122

Bruzzi, E., Talamini, F., Priftis, K., & Grassi, M. (2017). A SMARC effect for loudness. *I-Perception, 8*(6), 204166951774217. https://doi.org/10.1177/2041669517742175

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society b: Biological Sciences, 364*(1525), 1831–1840. https://doi.org/10.1098/rstb.2009.0028

Chang, S., & Cho, Y. S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00683

Cohen Kadosh, R., Brodsky, W., Levin, M., & Henik, A. (2008). Mental representation: What can pitch tell us about the distance effect? *Cortex, 44*(4), 470–477. https://doi.org/10.1016/j.cortex.2007.08.002

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Fairhurst, M. T., & Deroy, O. (2017). Testing the shared spatial representation of magnitude of auditory and visual intensity. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 629–637. https://doi.org/10.1037/xhp0000332

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Fernandez-Prieto, I., Spence, C., Pons, F., & Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *I-Perception, 8*(3), 204166951771618. https://doi.org/10.1177/2041669517716183

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research Psychologische Forschung, 65*(4), 250–259. https://doi.org/10.1007/s004260100065

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*(1), 95–110. https://doi.org/10.1080/135467996387552

Fischer, M. H., Riello, M., Giordano, B. L., & Rusconi, E. (2013). Singing numbers… in cognitive space—A dual-task study of the link between pitch, space, and numbers. *Topics in Cognitive Science, 5*(2), 354–366. https://doi.org/10.1111/tops.12017

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America, 5*(2), 82–108. https://doi.org/10.1121/1.1915637

Guilbert, A. (2020). About the existence of a horizontal mental pitch line in non-musicians. *Laterality, 25*(2), 215–228. https://doi.org/10.1080/1357650X.2019.1646756

Hartmann, M., & Mast, F. W. (2017). Loudness counts: Interactions between loudness, number magnitude, and space. *Quarterly Journal of Experimental Psychology, 70*(7), 1305–1322. https://doi.org/10.1080/17470218.2016.1182194

Heinemann, A., Pfister, R., & Janczyk, M. (2013). Manipulating number generation: Loud+long=large? *Consciousness and Cognition, 22*(4), 1332–1339. https://doi.org/10.1016/j.concog.2013.08.014

International Organization for Standardization. (2003). *Acoustics—Normal equal-loudness-level contours*. (ISO Standard No. 226:2003). https://www.iso.org/standard/34222.html

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32*(4), 662–673. https://doi.org/10.3758/BF03195857

Junghans, T., & Giblock, P. (2019). *LMMS: Linux MultiMedia studio*. https://lmms.io/

Koch, S., Schubert, T., & Blankenberger, S. (2023). The spatial representation of loudness in a timbre discrimination task. *I-Perception, 14*(6), 20416695231213212. https://doi.org/10.1177/204166952311213213

Lega, C., Cattaneo, Z., Ancona, N., Vecchi, T., & Rinaldi, L. (2020). Instrumental expertise and musical timbre modulate the spatial representation of pitch. *Quarterly Journal of Experimental Psychology, 73*(8), 1162–1172. https://doi.org/10.1177/1747021819897779

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 149–157. https://doi.org/10.1037/0278-7393.16.1.149

Macnamara, A., Keage, H. A. D., & Loetscher, T. (2018). Mapping of non-numerical domains on space: A systematic review and meta-analysis. *Experimental Brain Research, 236*(2), 335–346. https://doi.org/10.1007/s00221-017-5154-6

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics, 48*(2), 169–178. https://doi.org/10.3758/BF03207084

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16*(4), 406–419. https://doi.org/10.1037/a0024377

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Olejnik, S., Li, J., Supattathum, S., & Huberty, C. J. (1997). Multiple testing and statistical power with modified Bonferroni procedures. *Journal of Educational and Behavioral Statistics, 22*(4), 389–406. https://doi.org/10.3102/10769986022004389

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pitteri, M., Marchetti, M., Priftis, K., & Grassi, M. (2017). Naturally together: Pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians. *Psychological Research Psychologische Forschung, 81*(1), 243–254. https://doi.org/10.1007/s00426-015-0713-6

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ren, P., Nicholls, M. E. R., Ma, Y., & Chen, L. (2011). Size matters: Non-numerical magnitude affects the spatial coding of response. *PLoS ONE, 6*(8), e23553. https://doi.org/10.1371/journal.pone.0023553

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83*(2, Pt. 1), 274–278. https://doi.org/10.1037/h0028573

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimmean. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 297–338). Wiley.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99*(2), 113–129. https://doi.org/10.1016/j.cognition.2005.01.004

Schaal, N. K., Bauer, A.-K.R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe [The Gold-MSI: Replication and validation of a questionnaire for measuring musical sophistication in a german-speaking sample]. *Musicae Scientiae, 18*(4), 423–447. https://doi.org/10.1177/1029864914541851

Shaki, S., & Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition, 175*, 109–113. https://doi.org/10.1016/j.cognition.2018.02.022

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89*(4), 305–333. https://doi.org/10.1037/0033-295X.89.4.305

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *Afex: Analysis of factorial experiments*. https://CRAN.R-project.org/package=afex

Sixtus, E., Lonnemann, J., Fischer, M. H., & Werner, K. (2019). Mental number representations in 2D space. *Frontiers in Psychology, 10*, 172. https://doi.org/10.3389/fpsyg.2019.00172

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24*(1), 99–142. https://doi.org/10.1016/0010-0285(92)90004-L

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181. https://doi.org/10.1037/h0046162

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54*(6), 377–411. https://doi.org/10.1037/h0043680

Sutherland, C. A. M., Thut, G., & Romei, V. (2014). Hearing brighter: Changing in-depth visual perception through looming sounds. *Cognition, 132*(3), 312–323. https://doi.org/10.1016/j.cognition.2014.04.011

Takeshima, Y., & Gyoba, J. (2013). High-intensity sound increases the size of visually perceived objects. *Attention, Perception, & Psychophysics, 75*(3), 501–507. https://doi.org/10.3758/s13414-012-0403-z

The Audacity Team. (2019). *Audacity*. https://www.audacityteam.org/

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *The Journal of the Acoustical Society of America, 82*(4), 1193–1200. https://doi.org/10.1121/1.395255

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š, Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01798-5

Vellan, J. E., & Leth-Steensen, C. (2022). Separate processing mechanisms for spatial-numerical compatibility and numerical-size congruity. *Canadian Journal of Experimental Psychology/revue Canadienne De Psychologie Expérimentale, 76*(1), 44–56. https://doi.org/10.1037/cep0000270

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Walsh, V. (2015). A theory of magnitude: The parts that sum to number. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 552–565). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.64

Weis, T., Estner, B., Krick, C. M., Reith, W., & Lachmann, T. (2015). SNARC meets SPARC in fMRI—Interdependence of compatibility effects depends on semantic content. *Neuropsychologia, 77*, 331–338. https://doi.org/10.1016/j.neuropsychologia.2015.09.018

Weis, T., Estner, B., van Leeuwen, C., & Lachmann, T. (2016). SNARC (spatial-numerical association of response codes) meets SPARC (spatial-pitch association of response codes): Automaticity and interdependency in compatibility effects. *Quarterly Journal of Experimental Psychology, 69*(7), 1366–1383. https://doi.org/10.1080/17470218.2015.1082142

Weis, T., Theobald, S., Schmitt, A., van Leeuwen, C., & Lachmann, T. (2018). There's a SNARC in the size congruity task. *Frontiers in Psychology, 9*, 1978. https://doi.org/10.3389/fpsyg.2018.01978

Westfall, P., Johnson, W., & Utts, J. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika, 84*(2), 419–427. https://doi.org/10.1093/biomet/84.2.419

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience and Biobehavioral Reviews, 57*, 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science, 50*(4), 489–525.