BRIEF REPORT

The sound-free SMARC effect: The spatial-musical association of response codes using only sound imagery

Qi Jiang[1] · Atsunori Ariga[1]

Abstract

This study provides clear evidence that the human cognitive system automatically codes sound pitch spatially. The spatial-musical association of response codes (SMARC) effect, in which a high-pitched (low-pitched) tone facilitates an upper (lower) response, is considered to reflect the spatial coding of sound pitch. However, previous studies have not excluded the directional effects of sound localization. Because a high-pitched (low-pitched) tone is automatically misperceived as originating from a spatially high (low) location, the location of a perceived sound source might artificially elicit the SMARC effect. This study challenged this unresolved issue. Participants were trained to associate visual stimuli (novel contoured shapes) with sound pitches (high-pitched or low-pitched pure tones). After training, participants completed a discrimination task in which the vertically aligned keys were associated with the visual stimuli in the absence of sound. Even without sound, the SMARC effect was observed in response to the trained visual stimuli (Experiment 1). However, this sound-free SMARC effect was eliminated when training was omitted (Experiment 2). Therefore, the SMARC effect was observed based solely on the activation of sound imagery that was spatial.

Keywords SMARC effect · SPARC effect · Sound imagery

Introduction

The quickness and accuracy of an action toward a target object in the environment depend on the interaction between how it is encoded and how to respond to it. This phenomenon is evidenced by the stimulus-response compatibility (SRC) effect, which states that a right (left) response to a stimulus located on the right (left) is performed more quickly and accurately than when the opposite configuration is presented (Brebner, 1973; Fitts & Seeger, 1953; Kornblum, Hasbroucq, & Osman, 1990; Proctor & Vu, 2006; Shaffer, 1965; see also Hommel & Prinz, 1997, for a review). When an encoded stimulus has a consistent spatial code with the assigned response, the correct response is quickly selected due to dimensional overlap, whereas when this relationship is inconsistent, the response is disrupted by conflict between the spatial codes of the stimulus and those of the response (Kornblum et al., 1990; Kornblum & Lee, 1995; Proctor & Vu, 2006; Teichner & Krebs, 1974).

Although the locus of the SRC effect has been discussed in several previous studies (Proctor & Reeve, 1990), there is a consensus that this effect enables humans to understand how non-spatial stimuli are spatially encoded. For example, Westerners make a right (left) response quickly and accurately when a large (small) number is visually presented at the center of a display following the spatial representation of a mental number line oriented from left to right, although the number itself does not have spatial characteristics (Fias et al., 1996; Fischer et al., 2003). Interestingly, Rusconi et al. (2006) observed the SRC effect in response to sound pitch. Their participants were required to judge whether the pitch of the target sound (*E3*, *F3#*, *G3#*, *A3#*, *D4*, *E4*, *F4#*, or *G4#*) was higher or lower than that of a reference sound (*C4*) by pressing an upper (6) or lower (spacebar) key on a keyboard. Participants responded more quickly when using the upper key for high-pitched sounds and the lower key for low-pitched sounds than when the opposite configuration was presented. This phenomenon is known as either the spatial-musical association of response codes (SMARC) effect or spatial-pitch association

✉ Atsunori Ariga
  atsu.ariga@gmail.com

[1] Graduate School of Integrated Arts and Sciences, Hiroshima
  University, 1-7-1 Kagamiyama, Higashi-hiroshima,
  Hiroshima 739-8521, Japan

of response codes (SPARC) effect (Lidji et al., 2007). The SMARC effect is observed even when participants perform an instrumental timbre judgment task in which pitch height is irrelevant (Lidji et al., 2007); i.e., this effect automatically occurs even when attention is not focused on pitch height (Weis, Estner, & Lachmann, 2016a; Weis et al., 2016b). Therefore, the SMARC effect suggests that pitch height is coded spatially, which is the spatial-coding hypothesis.

However, it is possible that the SMARC effect can be explained by different mechanisms. High-pitched tones are misperceived as originating from spatially high locations and low-pitched tones are misperceived as originating from spatially low locations (Pratt, 1930; Roffler & Butler, 1968; Trimble, 1934). Considering this phenomenon, which is known as illusory sound localization, it can be assumed that the location of a perceived sound source rather than the spatial coding of sound pitch might artificially elicit the SMARC effect in response to the typical SRC; this is the illusory localization hypothesis. In fact, the location of a sound source is automatically detected and then exogenously cues a shift in visual attention (Spence & Driver, 1997) and the location (presented ear) of a task-irrelevant auditory stimulus automatically modulates lateralized motor performance (Nishimura & Yokosawa, 2009). Furthermore, recent studies that have assessed event-related potentials have reported that peripheral task-irrelevant auditory stimuli activate the visual cortex (McDonald et al., 2013), which suggests the automatic orientation of visual attention to sound location (Feng et al., 2014). Therefore, regarding the SMARC effect, the cognitive system may spontaneously and automatically localize (or mislocalize) a sound in the vertical dimension irrespective of its task relevance and thus prime vertically aligned responses based on the SRC. Indeed, whether this illusory sound localization is elicited by the spatial coding of sound pitch, the verbal metaphor of spatial positions for pitch height (e.g., "high" tone or "low" tone), or both remains unclear (for this discussion, see Ariga & Saito, 2019). However, it is necessary to demonstrate that the SMARC effect itself reflects our spatial coding of sound pitch, not the illusory sound localization, as current cumulative evidence for the spatial representation of pitch height relies heavily on this effect.

Ariga and Saito (2019) challenged the illusory localization hypothesis by demonstrating that the SMARC effect occurs in the absence of sound input. The authors examined whether the effect would be elicited by written pitch names alone and found that the SMARC effect occurred when participants judged pitch height labeled by visually presented word stimuli (C, D, E, G, A, and B). These participants also exhibited the effect in response to pitch names even when the indicated pitch height was irrelevant to the task. These results suggest that the

SMARC effect occurs at the semantic level in the absence of sound and excludes the directional effects of illusory sound localization on the corresponding response, which supports the spatial-coding hypothesis.

However, the findings of Ariga and Saito (2019) do not fully support the spatial-coding hypothesis because their stimuli were composed of typical notes (e.g., C, D, or E) which the participants may have already assigned to symbols (word labels). Based on this long-term learning, the participants could have automatically and internally read the visually presented word stimuli, retrieved tones with the corresponding pitch height, and easily depicted them on staves in their minds. That is, the visual spatial representation of staves rather than the spatial representation of pitch height might have been activated and elicited the SMARC effect.

To provide strong evidence demonstrating that pitch and space are associated via the spatial coding of pitch (spatial-coding hypothesis) rather than the mislocalization of sound (illusory localization hypothesis), the SMARC effect should be observed using sound imagery alone. Thus, this study examined whether the SMARC effect would occur following the presentation of stimuli that are difficult to depict on staves. More specifically, participants were trained to associate a new symbol system with the tones. Furthermore, atypical notes that could not be easily depicted on staves were used as stimuli. This approach gives us a better understanding of the nature of sound representation.

Experiment 1

Method

Ethics statement

All experimental procedures were reviewed and approved by the Institutional Review Board of Hiroshima University, Japan. Written informed consent was obtained from all participants before and after the experiment.

Participants

Based on a priori power analyses using G*Power (Faul et al., 2007), the present factorial design with a significance level of 0.05 and a power of 0.90, and a previously reported effect size ($d_Z = 0.7$; Ariga & Saito, 2019) for Experiment 1 (absence of sound), it was determined that a sample size of 24 would yield a 90.70% chance of correctly rejecting the null hypothesis. Thus, 24 Japanese participants (12 females, mean age = 21.63 years, SD = 1.63 years) were recruited from Hiroshima University; none were tone deaf or reported hearing or seeing problems and all were right-handed. Nine participants had no

experience reading or playing music whereas 15 had undergone musical training (0.5–19 years, mean period = 6.00 years, *SD* = 4.91 years) and were able to sight-read scores.

Stimuli

A preliminary experiment was conducted to select two neutral stimuli (*N* = 20, nine females, mean age = 21.20 years, SD = 1.25 years). Briefly, 100 novel contoured shapes (Endo, Saiki, Nakao, & Saito, 2003) were sequentially presented at the center of the display over a total of 100 trials. In each trial, one shape was presented twice: first, it appeared with a high pure tone (4,000 Hz) for 100 ms and then it appeared with a low pure tone (1,000 Hz) for 100 ms; there was an inter-stimulus interval of 1,000 ms; these two pitches are commonly used in hearing tests in Japan. Then, participants judged whether the presented stimulus shape was associated with a high or a low tone (two-alternative forced choice task). Based on the results, two neutral shapes (Patterns A and B; Fig. 1) that were not associated with a specific pitch height (i.e., judgment performance was at chance level) were selected as the target stimuli for the present study.

Procedure

The present experiment consisted of four sessions: two training sessions and two test sessions (see Type 1 sequence in Fig. 2).

Training session (Go/No-go task) Each training trial began with a fixation cross that appeared at the center of the display for 1,000 ms (Fig. 3). Then, one of the two target stimuli (Pattern A or B) was presented for 100 ms at the center of the display with a pure tone; Pattern A always accompanied a high tone and Pattern B always accompanied a low tone (see Training 1 for the Type 1 sequence in Fig. 2). Participants were required to quickly press "F" on a keyboard for Pattern A (go stimulus) but to suppress

Fig. 1 The two neutral target stimuli selected in the preliminary experiment

their response for Pattern B (no-go stimulus) and received feedback regarding the accuracy of the responses. The 500-ms feedback display was shown after a response or after 2,000 ms passed; thus, each participant had to make a response for the go stimulus within 2,000 ms after stimulus onset. This training session was performed to establish a new symbol system for pitch height (i.e., associations between novel visual stimuli and tones). Each participant performed 100 training trials in the Training 1 session.

Test session (discrimination task) Participants performed the test trials immediately after the training session. Each test trial began with a central fixation cross that appeared for 1,000 ms and was followed by one of the two target stimuli (Pattern A or B) in the absence of sound (Fig. 4), i.e., only visual stimuli were presented in the test session. Each participant was asked to determine whether the presented stimulus shape was Pattern A or B as quickly and accurately as possible by pressing "6" (upper key) on a keyboard for Pattern A or "spacebar" (lower key) for Pattern B (see Test 1 for the Type 1 sequence in Fig. 2). A response terminated the stimulus presentation and commenced the next trial; after each response, the 500-ms feedback display was shown. In the training session, Pattern A accompanied a high tone and Pattern B accompanied a low tone, and thus, this response mapping was congruent (congruent condition). Participants performed 100 test trials in the Test 1 session.

After the first test session, participants performed a second training session that was identical to the first training session (see Training 2 for the Type 1 sequence in Fig. 2). Subsequently, the participants performed a second test session that was identical to the first test session except that the response mapping was incongruent (see Test 2 for the Type 1 sequence in Fig. 2); i.e., the lower key was for Pattern A and the higher key was pressed for Pattern B (incongruent condition). This experiment included four sessions (two training sessions and two test sessions) with 100 trials each for a total of 400 trials. The order of the congruent and incongruent conditions was counterbalanced across participants; half performed the incongruent condition first by training with Pattern B as the go stimulus (see Type 2 sequence in Fig. 2).

One may argue that the present paradigm does not truly manipulate key alignment in the vertical direction (or a frontal plane) but, rather, that it manipulates the response keys in a transverse plane, relative to the participants. However, previous studies have shown that the SMARC effect is consistently observed, regardless of whether the response keys are aligned along the frontal plane (Pitteri et al., 2017) or transverse plane (Ariga & Saito, 2019; Lidji et al., 2007; Rusconi et al., 2006). Therefore, it is unlikely that key alignment influences the interpretation of the present results.
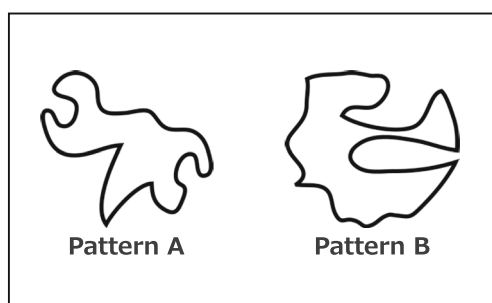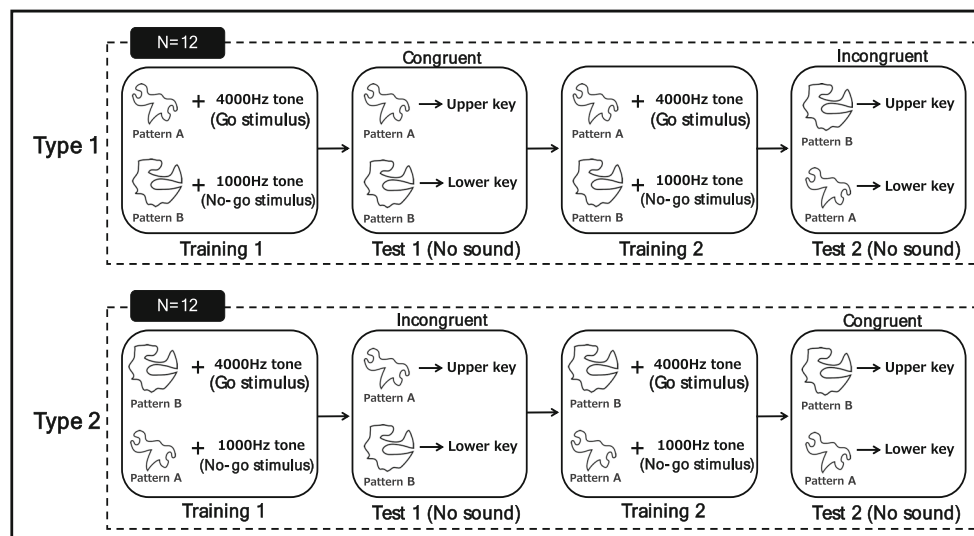
Fig. 2 Schematic illustration of Experiment 1

Results

The response times (RTs) on correct trials in the test sessions were averaged for each condition and each participant; the RTs for all participants are shown in Fig. 5. Paired t-tests revealed a significant difference between the congruent and incongruent conditions, $t(23) = 2.41$, $p = .024$, $d_Z = 0.31$. In addition, the error rates in the congruent and incongruent conditions were 3.71% and 4.25%, respectively, which indicates the absence of a speed-accuracy tradeoff.

Discussion

Performances in the test sessions were better when participants pressed the upper (lower) key for the visual stimulus that accompanied a high-pitched (low-pitched) tone in the training sessions (congruent condition) compared to the opposite configuration (incongruent condition). Importantly, the SMARC effect was triggered by the visual stimuli although the auditory stimuli were never presented in the test sessions. Because participants were trained with a new symbol system for the pitch heights of atypical notes, it was difficult for them to use a spatial representation based on staves. These results suggest that performance in the test trials was based on the activation of sound imagery that was evoked in response to the trained visual stimuli and elicited the sound-free SMARC effect. Consistent with the Ariga and Saito (2019), the present results clearly exclude the illusory localization hypothesis and strongly support the spatial-coding hypothesis.
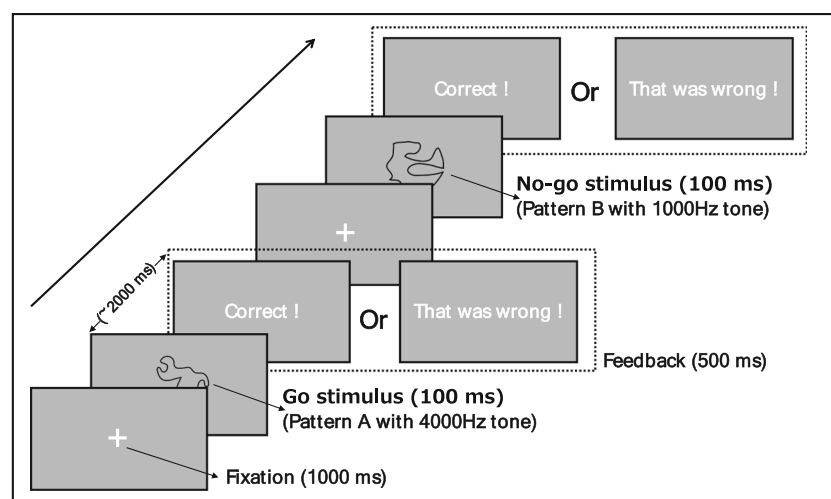

Fig. 3 Schematic illustration of the training trials for the Type 1 sequence in Experiment 1
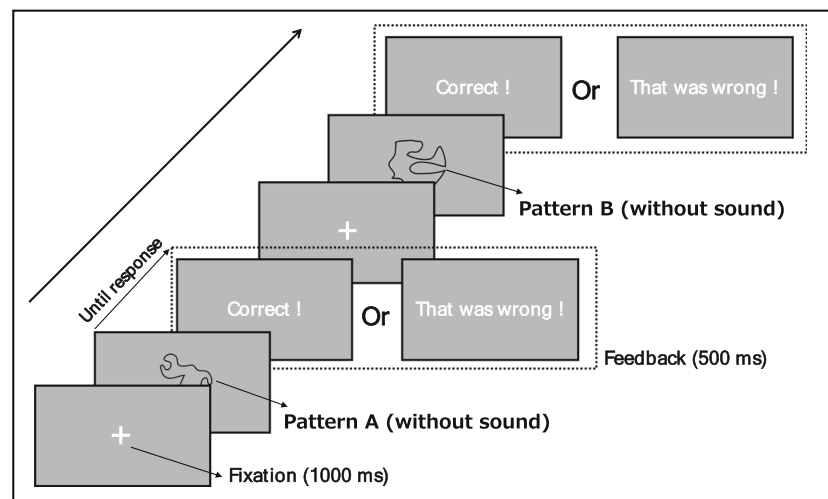
Fig. 4 Schematic illustration of the test trials in Experiment 1

Experiment 2 was performed to confirm the effect of the trained association between the visual patterns of the target shapes and pitch height. We predicted no SMARC effect when training was removed.

Experiment 2

Method

Participants

The sample size for Experiment 2 was determined as in Experiment 1; 24 Japanese participants (15 females, mean age = 19.50 years, SD = 1.15 years) were recruited from Hiroshima University. None were tone deaf or reported hearing or seeing problems and only one of the participants was left-handed.

Stimuli and procedure

The stimuli and procedure in Experiment 2 were identical to those used in Experiment 1 except that only test sessions were conducted. In the first session, participants were required to press the upper key for Pattern A and the lower key for Pattern B as quickly and as accurately as possible, in the same manner as participants in the Type 1 sequence in Experiment 1 were required to do in their first test session (simulated-congruent condition). In the second session, opposite mapping was used (simulated-incongruent condition). Experiment 2 consisted of two sessions with 100 trials each for a total of 200 trials. The order of conditions was counterbalanced across participants.

Results

The RTs for all participants are shown in Fig. 6. Paired t-tests revealed no significant difference between the two conditions, $t(23) = 0.79$, $p = .437$, $d_Z = 0.14$, and the error rates in the simulated-congruent and simulated-incongruent conditions were 3.25% and 2.96%, respectively.

To evaluate directly the effect of training, we also carried out cross-experiment 2 (training: presence in Experiment 1 and absence in Experiment 2) × 2 (congruency: congruence
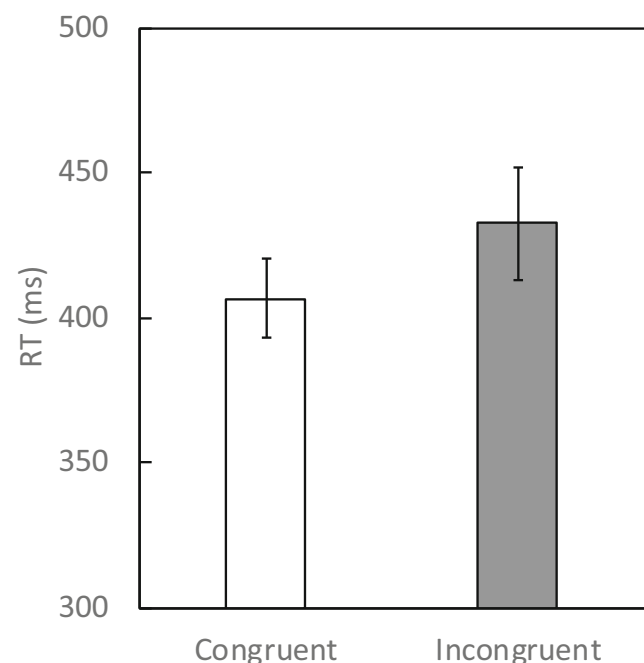

Fig. 5 Response times (RTs) under the congruent and incongruent conditions in the test sessions in Experiment 1. Error bars indicate the standard error of the mean
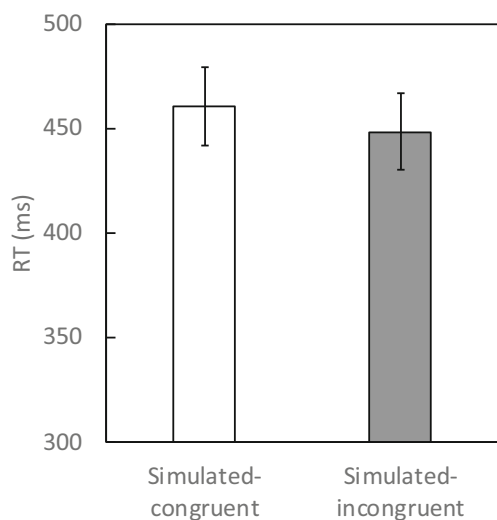
Fig. 6 Response times (RTs) under the simulated-congruent and simulated-incongruent conditions in Experiment 2. Error bars indicate the standard error of the mean

and incongruence) two-way analysis of variance on these data. It demonstrated a significant interaction, $F(1,46) = 4.031$, $p < .05$, partial $\eta^2 = 0.08$, which was derived from the significant effect of congruency in Experiment 1 and the non-significant effect in Experiment 2. The main effects of training, $F(1,46) = 2.213$, $p = .144$, partial $\eta^2 = 0.05$, and congruency, $F(1,46) = 0.486$, $p = .489$, partial $\eta^2 = 0.01$, were not significant.

Discussion

As expected, performance did not differ between the simulated-congruent and simulated-incongruent conditions when the training sessions were omitted. Therefore, the results in Experiment 1 were due to the training experience.

General discussion

In the present study, participants were trained on a new symbol system for pitch height and subsequently elicited the SMARC effect for visual stimuli (novel contoured shapes) associated with pitch height, even in the absence of sound input. This sound-free SMARC effect was likely not due to activated spatial representations of staves but to the sound imagery evoked by the trained association. Because the effect occurred using sound imagery without sound input, the present results suggest that pitch and space were associated via the spatial coding of pitch (spatial-coding hypothesis) rather than the mislocalization of sound (illusory localization hypothesis). These results advance the findings of Ariga and Saito (2019).

In the present study, the mean RT in the congruent condition was subtracted from that in the incongruent condition for

each participant in Experiment 1 and the difference was considered an index for the magnitude of the SMARC effect. The length of musical experience of the participants tended to be positively correlated with the magnitude of the SMARC effect ($r = 0.39$, $p = .06$), though this was a weak correlation and did not reach statistical significance. This result is consistent with previous reports showing that the SMARC effect is robustly strengthened by musical experience (Ariga & Saito, 2019; Lidji et al., 2007; Rusconi et al., 2006). Musical experience might enhance the strength and stability of the spatial coding of pitch height and/or increase the ability to activate a spatial code for pitch.

There is a possibility that the effect might be elicited by the spatial representation of loudness because we did not strictly control the perceived loudness of the tones (Bruzzi, Talamini, Priftis, & Grassi, 2017). Here, we predict that the loudness had little effect due to the small difference in the perceived loudness of the current auditory stimuli. However, further investigations are required to clarify this issue. The current findings suggest that the illusory sound localization is not responsible for the effect.

Previous studies have yet to fully elucidate the mechanisms underlying the SMARC effect, i.e., the spatial-coding hypothesis or illusory localization hypothesis. As a result, it remains unclear whether pitch and space are associated via the spatial coding of pitch height or through illusory sound localization. However, the present findings clearly demonstrate that the SMARC effect was due to the spatial coding of pitch height, which suggests that this effect occurred in response to sound imagery in the absence of sound input. The observation of a sound-free SMARC effect is consistent with the traditional and persisting view that pitch height is perceived in phenomenological space (Pratt, 1930).

The reason why sound pitch is described by adjectives that refer to space (e.g., "high" and "low" in English) in most languages of the world (Stumpf, 1883) has long been a mystery. The present findings suggest that this is because the coding of sound pitch is somehow spatial. Future research is needed to establish why it is spatial, as well as how it is related to the language we use to describe sound pitch.

Compliance with ethical standards

Declaration of conflicting interests None.

References

Ariga, A., & Saito, S. (2019). Spatial-musical association of response codes without sound. *Quarterly Journal of Experimental Psychology, 72*(9), 2288-2301.

Brebner, J. (1973). S-R compatibility and changes in RT with practice. *Acta Psychologica, 37*, 93-106.

Bruzzi, E., Talamini, F., Priftis, K., & Grassi, M. (2017). A SMARC Effect for Loudness. *i-Perception,8*(6), 204166951774217.

Endo, N., Saiki, J., Nakao, Y., &Saito, H. (2003). Perceptual judgments of novel contour shapes and hierarchical descriptions of geometrical properties. *Japanese Journal of Psychology, 74*(4), 346–353.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*,175-191.

Feng, W., Störmer, V. S., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2014). Sounds activate visual cortex and improve visual discrimination. *The Journal of Neuroscience, 34*, 9817-9824.

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*, 95-110.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience, 6*(6), 555-556.

Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 199-210.

Hommel, B., & Prinz, W. (1997). *Theoretical issues in stimulus response compatibility.* Amsterdam, The Netherlands: North-Holland.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility: A model and taxonomy. *Psychological Review,97*, 253–270.

Kornblum, S., & Lee, J. W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 855-875.

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*, 1189-1207

McDonald, J. J., Störmer, V. S., Martinez, A., Feng, W., & Hillyard, S. A. (2013). Salient sounds activate visual cortex automatically. *The Journal of Neuroscience, 33*, 9194-9201.

Nishimura, A., & Yokosawa, K. (2009). Effects of laterality and pitch height of an auditory accessory stimulus on horizontal response selection: The Simon effect and the SMARC effect. *Psychonomic Bulletin & Review, 16*, 666-670.

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, 81*, 243-254.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13*, 278-285.

Proctor, R. W., Reeve, T. G. (Eds.). (1990). Advances in psychology, Vol.65. *Stimulus-response compatibility: An integrated perspective.* Oxford, UK: North-Holland.

Proctor, R. W., & Vu, K. P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application.* Boca Raton, FL: CRC Press.

Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *Journal of the Acoustical Society of America, 43*, 1260-1266.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition, 99*, 113–129.

Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology, 70*, 284-288.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59*(1), 1-22.

Stumpf, C. (1883). *Tonpsychologie.* Leipzig, Germany: Verlag von S. Hirzel.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review, 81*, 75-98

Trimble, O. C. (1934). Localization of sound in the anterior posterior and vertical dimensions of auditory space. *British Journal of Psychology, 24*, 320-334

Weis, T., Estner, B., & Lachmann, T. (2016a). When speech enhances spatial musical association of response codes: Joint spatial association of pitch and timbre in nonmusicians. *The Quarterly Journal of Experimental Psychology, 69*.1689-1700.

Weis, T., Estner, B., van Leeuwen, C., & Lachmann, T. (2016b). SNARC meets SPARC: Automaticity and interdependency in compatibility effects. *The Quarterly Journal of Experimental Psychology, 69*, 1366-1383.