Language and Cognition (2023), 15: 4, 884–903
doi:10.1017/langcog.2023.31

ARTICLE

Sound-symbolic association between speech sound and spatial meaning in relation to the concepts of *up/down* and *above/below*

Lari Vainio[1,2], Alexandra Wikström[2], Claudia Repetto[3] and Martti Vainio[2]

[1]Perception, Action & Cognition Research Group, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland; [2]Phonetics and Speech Synthesis Research Group, Department of Digital Humanities, University of Helsinki, Helsinki, Finland; [3]Department of Psychology, Università Cattolica del Sacro Cuore, Milano, Italy
Corresponding author: Lari Vainio; Email: lari.vainio@helsinki.fi

(Received 13 February 2023; Revised 16 June 2023; Accepted 03 July 2023)

Abstract
Research has shown sound-symbolic associations between speech sounds and conceptual and/or perceptual properties of a referent. This study used the choice response time method to investigate hypothesized associations between a high/low vowel and spatial concepts of *up/down* and *above/below*. The participants were presented with a stimulus that moved either upward or downward (Experiments 1 and 2), or that was located above or below the reference stimulus (Experiment 3), and they had to pronounce a vowel ([i] or [æ]) based on the spatial location of the stimulus. The study showed that the high vowel [i] was produced faster in relation to the up-directed and the above-positioned stimulus, while the low vowel [æ] was produced faster in relation to the down-directed and the below-positioned stimulus. In addition, the study replicated the pitch-elevation effect showing a raising of the vocalization pitch when vocalizations were produced to the up-directed stimulus. The article discusses these effects in terms of the involvement of sensorimotor processes in representing spatial concepts.

Keywords: sound symbolism; spatial concepts; vocalization; speech

1. Introduction

Sound symbolism typically refers to the iconic associations that exist between speech sounds and conceptual and/or perceptual properties of a referent. These kinds of iconic connections between speech sounds and meaning are increasingly accepted as a general element of language (Perniss et al., 2010; Perniss & Vigliocco, 2014). Evolutionarily, these iconic connections might have claimed a place in language cognition by having a facilitatory role in language acquisition (Imai & Kita, 2014; Monaghan et al., 2012). The non-arbitrary association between speech sounds and meaning can be based on iconicity (i.e., form-meaning resemblance) and systematicity

(i.e., similar forms have similar meanings across words) (Dingemanse et al., 2015). In the present article, we use the term 'sound symbolism' solely in relation to an iconic relationship between speech sounds and meaning.

Research has identified sound-meaning associations in relation to various perceptual and conceptual properties such as emotion (Adelman et al., 2018; Körner & Rummer, 2022), motion speed (Cuskley, 2013), color (Anikin & Johansson, 2019; Johansson et al., 2020), and size (Sapir, 1929; Winter & Perlman, 2021). Most relevantly for the present study, research has revealed sound-symbolic links between speech sounds and spatial relations. For example, in the sound-distance effect, the back and low vowels have been associated with distal meaning referring to deictic concepts that are remote from an observer, while the front-high vowels have been associated with proximal meaning referring to deictic concepts that are near to an observer (Johansson & Zlatev, 2013; Rabaglia et al., 2016; Tanz, 1971; Vainio, 2021). Furthermore, Vainio et al. (2015, 2018) demonstrated another type of sound-space symbolism in which the front vowels [i] and [ø] were associated with forward-directed hand movements, while the back vowels [ɑ] and [o] were associated with backward-directed hand movements. Later investigations have revealed that this effect can be also observed in relation to forward/backward-directed leg movements suggesting that the effect might be based on the connection between processing front/back vowels and processing concepts of *forward/backward* (Vainio et al., 2019; Vainio & Vainio, 2021). Indeed, recent research revealed that the corresponding effect can be observed when the front ([i]) and back ([o]) vowel is produced based on whether the visual stimulus presents a front or back concept (Vainio et al., 2023), supporting the view that the effect originally observed in front- and back-directed limb movements reflects a general processing of *front* and *back* concepts. That is, the *front* concept is sound-symbolically associated with the vocalization of a front vowel, while the *back* concept is sound-symbolically associated with the vocalization of a back vowel. It was proposed that this sound-symbolic connection between the front/back vowel and the front/back concept might be based on associating semantic knowledge of *front* and *back* concepts to the articulatory front/back position of the tongue.

Regarding vertical space, Shintel et al. (2006) asked their participants to vocalize the sentence "*It is going up*" or "*It is going down*" depending on whether the visual stimulus moved up or down. The mean f_0 was found to be higher in relation to up vocalization in comparison to down vocalization. Similarly, Clark et al. (2013) have shown that participants use a higher pitch when reading stories related to high vertical space in comparison to low vertical space. It is not known, however, whether concepts linked to vertical space can be sound-symbolically associated with particular vowels. Hence, the present study investigates whether the spatial concepts of *up/down* and/or *above/below* could be associated with producing high/low vowels. The concept of *up/down* refers to the movement of something that is directed toward a higher/lower position, while the concept of *above/below* refers to a spatial position of something relative to a reference object. Our hypothesis was that high vowels would be associated with the concepts of *up* and *above*, while low vowels would be associated with the concepts of *down* and *below*. We had two reasons – that are not necessarily mutually exclusive – to assume that this would be the case. Firstly, high vowels have a relatively higher fundamental frequency (intrinsic vowel pitch; IVP) than low vowels (Sapir, 1989; Whalen & Levitt, 1995), and high/low tones are implicitly associated with high/low concepts (Pratt, 1930; Shintel et al., 2006; Spence, 2019), perhaps

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

because higher frequencies relatively frequently originate from elevated sources in the natural environment (Parise et al., 2014). As a consequence, this perceived resemblance might associate high/low vowels with the high/low vertical space. The second reason emphasizes sensorimotor compatibility between the perceived up- and down-directed movement of the visual stimulus and the up and down-directed movement of the larynx, tongue, and jaw required for executing high and low vowels, respectively (Fant, 1960; Honda et al., 1999; Ladefoged, 1968). This perspective is based on findings showing that some properties of the perceived stimulus – at least those that are potentially relevant for ongoing behavior – are automatically represented in action representations. As an example, the size of a viewed graspable object is automatically represented in grasp motor programs that correspond to the size (Franca et al., 2012; Tucker & Ellis, 2001), and perceiving the lip- or tongue-related phonemes selectively activates the lip- and tongue-related motor areas, respectively (Pulvermüller et al., 2006). As such, one might assume that perception of up- and down-directed stimuli is automatically represented in motor programs including the motor representations that are involved in producing up- (e.g., [i]) and down-directed (e.g., [æ]) articulations resulting in an association between particular vowels and the concepts of *up* and *down.*

Various experimental tasks have been used to investigate sound symbolism. Perhaps the most common methods in sound symbolism research are 2-alternative forced choice tasks (e.g., Margiotoudi & Pulvermüller, 2020; Nielsen & Rendall, 2012) and cross-linguistic methods (e.g., Blasi et al., 2016; Tanz, 1971). The present study uses the speeded choice reaction time (CRT) task in order to explore whether a particular speech sound is associated with a particular spatial concept. In CRT tasks, the overlap between stimuli and response is known to speed up performance (Kornblum et al., 1990). Most typically, this overlap occurs between spatial dimensions of visual stimuli and manual responses such as between the left–right location of the visual stimuli and the left–right hand as in the Simon effect (Simon, 1990). However, our previous research has shown that the CRT task, which measures reaction times of vocal responses, can be also used to investigate various sound symbolism phenomena (Vainio et al., 2017). For example, in Vainio's (2021) study participants were presented with target objects whose inter-object distance was either shortened or lengthened compared to the inter-object distance in the reference stimuli, and they were required to vocalize either [i] or [u] according to shortening/lengthening of the distance. It was found that vocalization responses were performed particularly rapidly in the hypothetically congruent block (i.e., [i]-short, [u]-long) in comparison to the incongruent block. Similarly, in the present study, we investigate whether a potential overlap between the spatial dimension of the visual stimuli and the vocalization response can be observed in reduced reaction times of vocal responses. Hence, the participants are asked to respond to the stimuli in hypothetically congruent and incongruent conditions. In the congruent condition, they were asked to vocalize [i] for the upward-directed movement and [æ] for the downward-directed movement, and vice versa in the incongruent condition. We propose that if the concepts of *up/above* and *down/below* are sound-symbolically associated with high/low vowels, the responses should be produced particularly rapidly in congruent conditions in comparison to incongruent conditions.

In addition to measuring response times of vocalizations, we also measure the vocal characteristics of f_0 (fundamental frequency), *F1* (first formant), and *F2* (second formant). It is known that the *F1* of vocalization increases with the opening of the oral

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

cavity (Fant, 1960). Hence, low vowels have typically higher *F1* values than high vowels. In addition, the *F2* is known to increase when vocalizations are produced with the increased front position of the tongue. Consequently, front vowels have typically higher *F2* values than back vowels. Finally, as mentioned above, the f_0 is known to be increased with high in comparison to low vowels. Regarding these vocal characteristics, it could be hypothesized, for example, that if the perceived upward-directed movement is indeed associated with high vowels, the upward-directed stimuli could decrease the *F1* values in comparison to the downward-directed stimuli. Furthermore, it is expected that similar to that reported by Shintel et al. (2006), f_0 values might be increased when vocalizations are produced to the upward-directed stimuli in comparison to the downward-directed stimuli.

The study consists of three experiments that investigate whether high vowel [i] would be associated with the concepts *up* and *above* and whether low vowel [æ] would be associated with the concepts *down* and *below*. The first experiment investigates whether the production of the unrounded high-front vowel [i] is associated with the upward-directed movement of the stimuli and whether the production of the unrounded low-front vowel [æ] is associated with the downward-directed movement of the stimuli. Basically, due to our selection criteria, we were not left with the possibility to use other vowels than [i] and [æ]. That is, we had three reasons for selecting these vowels. Firstly, they are suitable for our hypothesis requiring a high–low contrast between the response alternatives. Secondly, they are not included in the words *up, down, above,* or *below* in the languages that were mastered by the participants (see the Participants section below). Thirdly, they presented phonemes that were allowed by Finnish phonology and phonotactics[1]. The second experiment primarily provides a control study for the first experiment using different speech sounds to those used in the first experiment. Finally, the third experiment replicates the first experiment with the exception that instead of exploring sound symbolism in relation to the concepts of *up* and *down,* the experiment explores sound symbolism in relation to the concepts of *above* and *below*.

2. Experiments 1 and 2

Experiment 1 investigates whether the vowel [i] is produced faster when the target stimulus moves upward in comparison to downward and whether the vowel [æ] is produced faster when the target stimulus moves downward in comparison to upward. In addition, the experiment explores whether the vocal characteristics of f_0, *F1*, and/or *F2* could be modulated by the up–down direction of the stimulus. Based on the previous findings (Shintel et al., 2006), it was expected that at least f_0 values could be modulated by up/down-directed stimuli so that f_0 values would be increased when vocal responses are performed to up-directed stimuli.

Experiment 2 provides a control study for Experiment 1. It should be noticed that the letter *i* is a much more commonly used letter in the Finnish language than the letter ä (i.e., the phoneme [æ]) (Pääkkönen, 1991). Therefore, based on the linguistic markedness account (Zimmer, 1964), it can be assumed that, due to being more frequent, the [i] is a non-marked item, while the [æ] is a marked item. Similarly, it has been proposed that the spatial dimension of up is non-marked, while down is marked

[1]The Finnish vowel system consists of the following vowels: [ɑ], [e], [i], [o], [u], [y], [æ], and [ø].

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

because the word up is more frequent than down (Winter et al., 2015). As such, it is possible that the congruency effect observed in Experiment 1 presents a version of the so-called MARC (Markedness Association of Response Codes) effect (Lakens, 2012; Willmes & Iversen, 1995) in which the congruency between the non-markedness/markedness of the response ([i]/[æ]) and the non-markedness/markedness of the stimulus (up/down) causes the observed congruency effect. If this were the case, we should observe that, in addition to [i], the speech sound [v] should be associated with faster responses when the stimulus moves upward, and, in addition to [æ], the speech sound [f] should be associated with faster responses when the stimulus moves downward. That is, because, due to being more frequent, [v] is a non-marked item, while [f] is a marked item (Pääkkönen, 1991). Hence, Experiment 2 replicates Experiment 1 with the exception that the responses [i] and [æ] will be replaced by the responses [fe] and [ve]. The reason why we used the consonant–vowel structure in responses rather than asking participants to solely pronounce the consonants was that it is very difficult to pronounce [v] in the absence of any surrounding vowel.

2.1. Methods

2.1.1. Participants

Twenty-four volunteers naïve to the purposes of the experiment participated in Experiment 1 (19–40 years of age; mean age = 27 years; 5 males; 1 left-handed), and twenty-five volunteers naïve to the purposes of the experiment participated in Experiment 2 (19–44 years of age; mean age = 27.3 years; 1 male; 2 left-handed). All participants were native speakers of Finnish and reported normal hearing and normal or corrected-to-normal vision. In both Experiments, all participants mastered Finnish, English, and Swedish. In Experiment 1, one participant mastered Spanish, one mastered Estonian, and one mastered German. In Experiment 2, one participant mastered Estonian. The participants did not master any other languages. Power was estimated based on simulations (Brysbaert & Stevens, 2018). The simulations were based on an earlier dataset from an experiment with a very similar design (Vainio et al., 2023). In the simulations, a mixed linear model with log-transformed reaction time data was fitted. The participants had a random effect on the intercept and the slope of congruency. The simulations suggest, firstly, that with the effect size (d_z = 0.42) observed by Vainio et al. (2023), 22 observers would have sufficed to produce a statistically significant difference in 85% of experiments. The simulations were run with R package simr (Green & MacLeod, 2016). All of the participants reported being unaware of the purpose of the study and the nature of the investigated effect. Written informed consent was obtained from all participants. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioral Sciences at the University of Helsinki.

2.1.2. Stimuli, procedure, and apparatus

Each participant sat in a dimly lit room with his or her head 75 cm in front of a 19″ CRT monitor (screen refresh rate: 85 Hz; screen resolution: 1280 × 1024). A head-mounted microphone was adjusted close to the participant's mouth. At the beginning of each trial, a blank white screen was presented for 2,700 ms. Then the reference stimulus (Fig. 1: frame-b) was presented for 1,000 ms. The size of the stimulus was 7.6° (horizontally) x 5° (vertically). After that, the target stimulus, which was the same

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

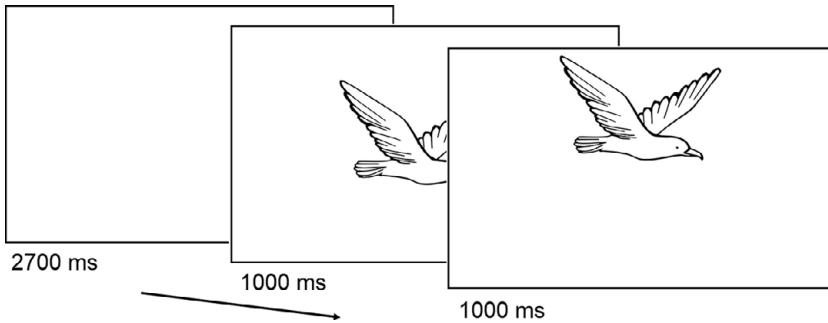

Figure 1. An illustration of the design used in Experiments 1 and 2.

bird as in the reference stimulus (Fig. 1: frame-c), was presented for 1,000 ms. In the target stimulus, the bird was presented either higher or lower (1.9°) than the reference stimulus. This design is illustrated in Fig. 1. In Experiment 1, participants were required to pronounce a single vowel ([i] or [æ]) according to the up/down location of the target. The vowels [i] and [æ] were selected for the study firstly because [i] as a high vowel and [æ] as a low vowel make them suitable for exploring whether spatial compatibility between response (high vowel versus low vowel) and stimulus (high location versus low location) results in a congruency effect, and secondly, because these sounds are not dominant in the words *up/down* in the languages that were mastered by the participants (i.e., Finnish: /yløs/—/ˈɑlɑs/; Swedish: /ʊp/—/neːd/; Spanish: /aˈriβa/—/abaˈxo/; Estonian: /ˈyles/—/ˈɑlːɑ/; German: /hɛˈraʊ̯f/—/hɛˈrʊntɐ/). In Experiment 2, participants were required to pronounce a single syllable ([fe] or [ve]) according to the up/down location of the target.

Experiments 1 and 2 were divided into two blocks that were separated by a 10-minute break. Each block lasted for approximately 10 minutes. In one block (i.e., the congruent block), the participants were required to vocalize [i] (or [fe] in Experiment 2) if the target bird moved upward and [æ] (or [ve] in Experiment 2) if it moved downward. In the other block (i.e., the incongruent block), the participants were required to vocalize [æ] (or [ve] in Experiment 2) if the bird moved upward and [i] (or [fe] in Experiment 2) if it moved downward. The beginning of each block included a practice session that presented 20 repetitions including the same number of hypothetically congruent and incongruent stimulus–response conditions.

Half of the participants performed first the [i]-up (or [fe]-up in Experiment 2) block. Both blocks consisted of 15 stimulus conditions in which the face of the bird was pointing to the left and it moved upward, 15 stimulus conditions in which the face of the bird was pointing to the left and it moved downward, 15 stimulus conditions in which the face of the bird was pointing to the right and it moved upward, and 15 stimulus conditions in which the face of the bird was pointing to the right and it moved downward. All stimuli were presented in randomized order and on a white background. In total, both experiments consisted of 120 trials [30 repetitions × 2 (block) × 2 (direction)].

The participants were instructed to pronounce the vowel/syllable as quickly as possible after the onset of the target stimulus. It was emphasized that the vowel should be uttered as a short (e.g., [i]) rather than a long (e.g., [iː]) vowel in a natural talking voice. During the practice session, the experimenter verified that the vocalizations

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

met these criteria. After Experiment 1, the participants were asked whether the task was easier to perform in the [i]-upward/[æ]-downward or in the [i]-downward/[æ]-upward block. After Experiment 2, the participants were asked whether the task was easier to perform in the [fe]-up/[ve]-down or in the [fe]-down/[ve]-up block. In addition, in both experiments, they were asked whether they noticed that they would have mirrored the up/down movement of the bird by raising or lowering their vocalization pitch.

Sound recording and stimulus presentation were carried out with Presentation® software (Version 16.1, www.neurobs.com). The vocal responses were recorded for 2,000 ms starting from the onset of the target object. At the beginning of the experiment, the recording levels were calibrated for each participant using the voice calibration function of Presentation® software so that the recording levels would match the natural intensity of the participant's voice.

2.1.3. Statistical analyses

For analysis purposes, the onsets of the vocalizations were located for each trial as the first observable peak in the acoustic signal. Similarly, the offsets of the vocalization were located individually for each trial as the observable ending of the acoustic signal. For this task, onsets and offsets were initially located by a highly experienced person who was blind to the condition of each acoustic signal. The spectral components (F_1 and F_2), as well as $f_{0,}$ were calculated as median values of the middle third of the voiced section of the vowel. The procedure of locating the onsets and offsets of vocalizations, as well as calculating the acoustic parameters, was carried out in the same way as in our previous studies that similarly investigate reaction times of vocalizations (e.g., Vainio, 2021; Vainio et al., 2017).

The following parameters were analyzed from the raw data: reaction times, f_0, *F1,* and *F2.* On a few occasions, the formant value was not found by Praat (Version 6.2.15; Boersma, 2001) or the output value clearly exceeded variations that can normally be observed within the voice characteristics of the given vowels (e.g., octave jump errors) (Experiment 1: f_0: 1.5%; *F1*: 5.8%; *F2*: 4.2%; Experiment 2: f_0: 1.7%; *F1*: 0.7%; *F2*: 0.5%). The particularly large number of missing values of F1 and F2 are due to breathy voice quality as some participants produced vocalizations rather quietly. Such values, as well as values that were more or less than two standard deviations from a participant's median (Experiment 1: f_0: 0.4%; *F1*: 1.7%; *F2*: 0.9%; Experiment 2: f_0: 1.1%; *F1*: 0.2%; *F2*: 0.2%), were discarded prior to analyzing the acoustic characteristics of the vocalizations. However, prior to analyzing any of these parameters, the errors (i.e., the participant uttered the wrong speech unit or did not produce any response) were removed from the data (Experiment 1: 0.6%; Experiment 2: 1.6%). In the analysis of reaction times, reaction times faster than 250 ms (Experiment 1: 0.3%; Experiment 2: 0.4%) and slower than 1,000 ms (Experiment 1: 0.07%; Experiment 2: 1.1%) were excluded from the analysis because it has been shown that responding to a visually presented target takes a minimum of 200–300 ms (Welford, 1980). In Experiment 2, the data of one participant was removed from the analyses because 15% of her responses were either missing (12%) or incorrect (3%) (see Bennett, 2001). In addition, for analyzing fundamental frequencies, the raw f_0 values were converted to semitone (*st*) movements relative to each participant's mean f_0. Semitone conversion was conducted to account for the logarithmic nature of the perceiving pitch and pitch movements and to eliminate the bimodal distribution of

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

fundamental frequencies caused by the fact that male speakers have fundamentally lower f_0 values than female speakers.

The output of the normality test indicated that the data were not normally distributed in Experiments 1 and 2. Consequently, the statistical significance of observed differences was tested using the generalized linear mixed model analysis (GLMM) with gamma distribution assumption (log link function). The GLMM analysis treated Location (1 = up; 2 = down) and Response [1 = [i] (or [fe] in Experiment 2); 2 = [æ] (or [ve] in Experiment 2)] as fixed within factors. The subject was allowed to have a random effect on the intercept and the slope of Location and Response. All pairwise comparisons were carried out using Bonferroni correction for multiple comparisons. The analysis was carried out using the SPSS statistics software package (version 28). For effect sizes, standardized effect sizes (Cohen's d_z; see Lakens, 2013) were calculated. General guidelines for d are small (>0.2), medium (>0.5), and large (>0.8).

2.2. Results

In Experiment 1, the analysis of reaction times revealed a significant interaction between Location and Response [$F(1,2849) = 91.45$, $p < .001$]. The pairwise comparisons test showed that [i] responses were performed faster when the target was presented in the up location (M = 432 ms) rather than the down location (M = 454 ms) ($p < .001$, $d_z = 0.25$), while [æ] responses were performed faster when the target was presented in the down location (M = 434 ms) rather than the up location (M = 464 ms) ($p < .001$, $d_z = 0.33$). These observations are presented in Fig. 2. In Experiment 2, the analysis of reaction times showed that the interaction [$F(1,2801) = 2.66$, $p = .103$] was not significant. None of the main effects were significant. These observations are presented in Fig. 3.

2.2.1. Cross-experiment analysis

We analyzed whether the congruency effect observed particularly in reaction times of Experiment 1 differs significantly between Experiments 1 and 2 by extracting Location and Response data into one Congruency variable in which congruent

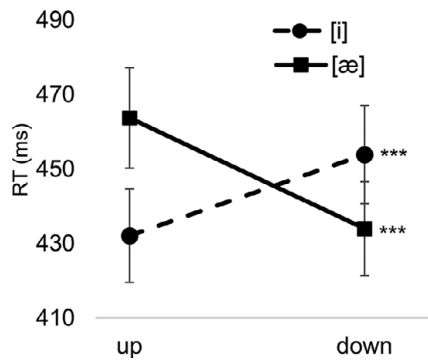

Figure 2. The mean vocal reaction times for Experiment 1 as a function of the response (vowel) and the up/down location of the target stimulus. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences (***$p < .001$).

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

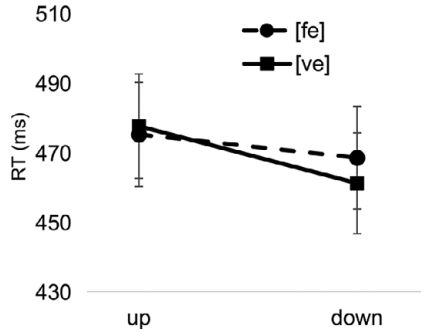

Figure 3. The mean vocal reaction times for Experiment 2 as a function of the response (syllable) and the up/down location of the target stimulus. Error bars depict the standard error of the mean.

responses referred to conditions in which Target location was up and Response was [i]/[fe] and Location was down, and Response was [æ]/[ve]. Opposite Location-Response mapping was applied for incongruent responses. A significant interaction between Congruency and Experiment [$F(1,5654) = 25.15$, $p < .001$] was observed in this analysis. This shows that the congruency effect between Location and Response significantly differs between Experiments 1 and 2. The congruency effect is highly significant in Experiment 1 ($p < .001$), while in Experiment 2 the effect is missing ($p = .125$).

In Experiment 1, the analysis of f_0 values revealed significant main effects of Location [$F(1,1392) = 11.96$, $p < .001$] ([up]: $M = 0.67$ st, [down]: $M = 0.51$ st, $d_z = 0.43$) and Response [$F(1,1392) = 22.80$, $p < .001$] ([i]: $M = 0.72$ st, [æ]: $M = 0.48$ Hz, $d_z = 0.39$) as well as the interaction between these two factors [$F(1,1392) = 10.64$, $p < .001$]. The pairwise comparisons test showed that for [i] responses the st values were significantly higher in the up location (M = 0.86 st) than in the down location (M = 0.58 st) ($p < .001$; $d_z = 0.66$). However, this effect was not significant for [æ] responses (p = .383). Regarding the analysis of *F1* values, the main effect of Response was significant [$F(1,2644) = 98.13$, $p < .001$], showing that [æ] responses were higher (M = 592 Hz) than [i] responses (M = 313 Hz) ($d_z = 0.98$). Regarding F2 values, the [i] responses were significantly [$F(1,2711) = 782.20$, $p < .001$] higher (M = 2855 Hz) than [æ] responses (M = 1490 Hz) ($d_z = 4.2$).

In Experiment 2, the analysis of f_0 values showed significant main effects of Location [$F(1,1317) = 5.79$, $p = .016$] ([up]: $M = 0.66$ st, [down]: $M = 0.54$ st, $d_z = 0.31$) and Response [$F(1,1317) = 4.45$, $p = .035$] ([fe]: $M = 0.66$ st, [ve]: $M = 0.55$ st, $d_z = 0.27$). The interaction was not significant [F(1,1317) = 3.80, p = .052]. Regarding the analysis of *F1* values, the main effect of Response [$F(1,2816) = 28.79$, $p < .001$] ([fe]: $M = 530$ Hz; [ve]: $M = 500$ Hz, $d_z = 0.24$) was significant. The analysis of *F2* values revealed a significant main effect of Response [$F(1,2824) = 6.73$, $p = .010$] ([fe]: $M = 2199$ Hz; [ve]: $M = 2173$ Hz, $d_z = 0.11$). However, notice that the effect size is very small.

Finally, in Experiment 1, 15 participants (62.5%) judged that the task was easier to perform in the [i]-upward/[æ]-downward block, and 5 participants (20.8%) judged that the task was easier to perform in the [i]-downward/[æ]-upward block. The rest of the participants did not report any difference between the blocks. Two participants mentioned that they might have involuntarily raised/lowered their vocalization pitch

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

according to the up/down movement of the bird. In Experiment 2, 10 participants (41.6%) judged that the task was easier to perform in the [fe]-upward/[ve]-downward block, and 5 participants (20.8%) judged that the task was easier to perform in the [fe]-downward/[ve]-upward block. The rest of the participants did not report any difference between the blocks. Four participants mentioned that they might have involuntarily raised/lowered their vocalization pitch according to the up/down movement of the bird.

2.3. Discussion

The results of Experiment 1 revealed the stimulus–response compatibility effect between the up/down location of the target stimulus and the high/low response. The high vowel [i] was produced particularly rapidly when the target was presented at the up location, while the low vowel [æ] was produced particularly rapidly when the target was presented at the down location. Considering the analysis of the *F1* or *F2,* the results did not reveal any other effects than the expected difference in *F1* and *F2* values between the two responses. That is, *F1* is larger for [æ] than [i], and *F2* is larger for [i] than [æ]. However, the analysis of the f_0 values showed, firstly, that the f_0 is significantly higher for the [i] responses than for the [æ] responses replicating the IVP phenomenon. In addition, f_0 was significantly higher when the target was presented at the up location in comparison to the down location. This was particularly the case in relation to [i] and [e] (i.e., [fe] and [ve]) responses. This effect was not observed for [æ] responses in Experiment 1. This finding is somewhat in line with the previous observations (e.g., Shintel et al., 2006) presenting that people show an implicit tendency to produce higher-pitched vocalizations when the vocalizations are produced to up-directed stimuli in comparison to down-directed stimuli. However, the fact that the effect was not observed in relation to [æ] suggests vowel articulation processes can modify this effect so that the effect is diminished with those vowel articulations whose IVP is naturally relatively low.

The results of Experiment 2 did not replicate the congruency effect that was observed in Experiment 1. The [fe] or [ve] responses were not associated with facilitated responses when the target was presented at the up or down location. The analysis showed that the congruency effect significantly differs between Experiment 1 and Experiment 2; in Experiment 2 the effect was missing. This finding supports the view that the congruency effect observed in the response times of Experiment 1 is not likely to be based on the markedness mapping processes. Consequently, the account according to which the congruency effect observed in Experiment 1 reflects sound-symbolic associations between speech sounds of [i]/[æ] and spatial concepts of up/down remains a plausible explanation of the effect.

3. Experiment 3

Experiment 3 investigates whether the congruency effect observed in the response times of Experiment 1 can be replicated when the concept of *up/down* is replaced by the concept of *above/below*. For this purpose, the vowels [i] and [æ] are pronounced according to the spatial relationship between the target (a bird) and the reference (a cloud). Hence, instead of responding to the up/down-directed movement of the

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

target bird as in Experiment 1, in this experiment, the participants are asked to pronounce [i]/[æ] if the bird is above/below the cloud.

3.1. Methods

3.1.1. Participants

Twenty-three volunteers naïve to the purposes of the experiment participated in Experiment 3 (20–38 years of age; mean age = 25.8 years; 4 male; 4 left-handed). All participants were native speakers of Finnish and reported normal hearing and normal or corrected-to-normal vision. All participants mastered Finnish, English, and Swedish, and one participant mastered German. All of the participants reported being unaware of the purpose of the study and the nature of the investigated effect. Written informed consent was obtained from all participants. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioral Sciences at the University of Helsinki.

3.1.2. Stimuli, procedure, and apparatus

The apparatus, environmental conditions, and calibration were the same as those in Experiments 1 and 2. The procedure and design were also similar to that of Experiment 1 with the exception that instead of presenting participants with a bird that moved up or down, the stimuli consisted of a bird (6.1° horizontally, 3.4° vertically) that was located above or below a cloud (10.7° horizontally, 3.8° vertically). The distance between the bird and the cloud was 0.8°. At the beginning of each trial, a blank white screen was presented for 2,500 ms. Then the target object was presented for 1,400 ms. The participants were required to vocalize the vowel [i] or [æ] depending on whether the bird was located above or under the cloud. The face of the bird was pointing to the right in each stimulus. There were two 'above' conditions: the cloud was located at the center of the display (above-high), or the bird was located at the center of the display (above-low). In addition, there were two 'below' conditions: the bird was located at the center of the display (below-high), or the cloud was located at the center of the display (below-low). The vowels [i] and [æ] were suitable for exploring whether spatial compatibility between response (high vowel versus low vowel) and stimulus (above versus below) results in a congruency effect because these sounds are not included in the words *above/below* in the languages that were mastered by the participants (respectively; Finnish: /ˈylæˌpuo̯elːɑ/—/ˈɑlɑˌpuo̯elːɑ/; Swedish: /ˈøːvɛr/—/ˈʊndɛr/; German: /ˈyːbɐ/—/ˈʊntɐ/). As can be noticed, the vowel [æ] occurs in the word *above* in Finnish (/ˈylæˌpuo̯elːɑ/). However, as it was hypothesized that this vowel should be associated with the concept of *below* rather than *above*, we concluded that this overlap did not cause any theoretical problem. That is, the overlap between the vowel [æ] and the word /ˈylæˌpuo̯elːɑ/ should result in a negative congruency effect instead of a positive congruency effect, which works against the hypothesis, making the potential interaction effect even more convincing.

Half of the participants performed first the [i]-above block. Both blocks consisted of 15 stimulus conditions in which the bird was above the cloud and the cloud was located at the center of the display (1: higher above condition), 15 stimulus conditions in which the bird was above the cloud and the bird was located at the center of the display (2: lower above condition), 15 stimulus conditions in which the bird was below the cloud and the bird was located at the center of the display (3: higher below

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

condition), and 15 stimulus conditions in which the bird was below the cloud and the cloud was located at the center of the display (4: lower below condition) (see Fig. 4 for these stimulus conditions).

The voice calibration, sound recording, stimulus presentation, Praat analysis, and semitone conversion were carried out similarly to that in Experiments 1 and 2. After Experiment 3, the participants were asked whether the task was easier to perform in the [i]-above/[æ]-below or in the [i]-below/[æ]-above block. In addition, they were asked whether they noticed that they would have mirrored the above/below location of the bird by raising or lowering their vocalization pitch. The beginning of each block included a practice session that presented 20 repetitions including the same number of hypothetically congruent and incongruent stimulus–response conditions. In total, Experiment 3 consisted of 120 trials [15 repetitions × 2 (block) × 2 (above/below position) × 2 (bird located at the center/cloud located at the center)].

3.1.3. Statistical analyses

The following parameters were analyzed from the raw data: reaction times, f_0, *F1*, and *F2* (*Table 1*). Regarding the voice characteristics, the same criteria were employed for removing outliers (f_0: 0.9%; *F1*: 5.5%; *F2*: 2.4%) as in Experiments 1 and 2. Prior to analyzing any of these parameters, the errors (i.e., the participant uttered the wrong speech unit or did not produce any response) were removed from the data (0.5%). In the analysis of reaction times, reaction times faster than 250 ms (0.0%) and slower than 1,000 ms (1.8%) were excluded from the analysis before carrying out the reaction time analysis.

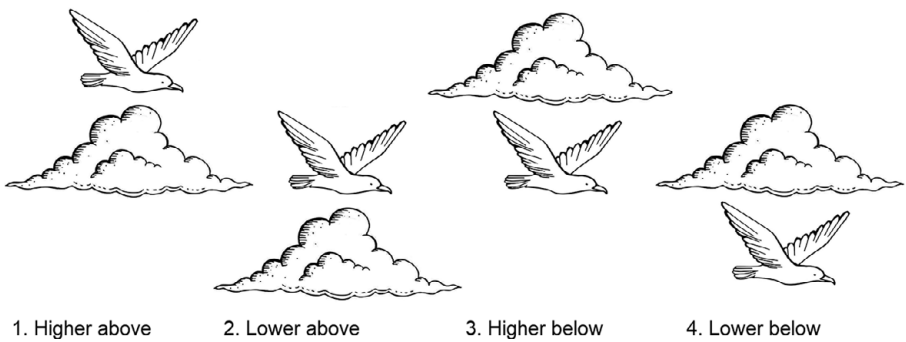

Figure 4. The stimuli used in Experiment 3.

Table 1. The means and SEs of reaction times, f_0s, *F1*s, and *F2*s of Experiment 3 in all possible Stimulus and Response combinations

Stimulus	Response	mean RT (ms)	f_0 (st)	*F1* (Hz)	*F2* (Hz)
higher – above	[i]	575 (13.9)	0.69 (0.07)	312 (16.6)	2,857 (39.4)
	[æ]	610 (14.8)	0.59 (0.08)	532 (28.4)	1,413 (19.5)
lower – above	[i]	597 (14.5)	0.70 (0.07)	313 (16.6)	2,866 (39.4)
	[æ]	625 (15.1)	0.50 (0.06)	531 (28.3)	1,418 (19.6)
higher – below	[i]	614 (14.9)	0.64 (0.07)	301 (16.0)	2,863 (39.4)
	[æ]	598 (14.5)	0.48 (0.06)	558 (29.7)	1,444 (19.9)
lower – below	[i]	610 (14.8)	0.54 (0.06)	300 (15.9)	2,868 (39.5)
	[æ]	584 (14.1)	0.51 (0.06)	535 (28.5)	1,437 (19.8)

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

The statistical significance of observed differences was tested using GLMM with gamma distribution assumption (log link function). The GLMM analysis treated Location (1 = above; 2 = below), Height [1 = higher above/below conditions (images 1 and 3 in Fig. 4); 2 = lower above/below conditions (images 2 and 4 in Fig. 4)], and Response (1 = [i]; 2 = [æ]) as fixed within factors. The subject was allowed to have a random effect on the intercept and the slope of Location, Height, and Response. All pairwise comparisons were carried out using Bonferroni correction for multiple comparisons. The analysis was carried out using the SPSS statistics software package (version 28).

3.2. Results

In the analysis of reaction times, we observed significant interactions between Location and Response [$F(1,2689) = 59.55$, $p < .001$] as well as Location and Height [$F(1,2689) = 16.61$, $p < .001$]. None of the main effects were significant. In addition, the two-way interaction between Height and Response [$F(1,2689) = 1.69$, $p = .193$] as well as the three-way interaction between Location, Height, and Response [$F(1,2689) = 0.09$, $p = .771$] were not significant. According to the pairwise comparisons test, the [i] responses were performed faster when Location was above ($M = 586$ ms) rather than below ($M = 612$ ms) ($p < .001$, $d_z = 0.29$), while the [æ] responses were performed faster when Location was below ($M = 591$ ms) rather than above ($M = 617$ ms) ($p < .001$, $d_z = 0.29$). These observations are presented in Fig. 5. Furthermore, the above condition was associated with faster responses in the higher condition ($M = 592$ ms) rather than the lower condition ($M = 611$ ms) ($p < .001$, $d_z = 0.19$), while the below condition was associated with somewhat faster responses in the lower condition ($M = 597$ ms) rather than the higher condition ($M = 606$ ms), although this effect was not significant ($p = .065$, $d_z = 0.10$).

Considering the analysis of f_0 values, a main effect of Response [$F(1,1313) = 9.97$, $p = .002$] ([i]: $M = 0.64$ st, [æ]: $M = 0.52$ st, $d_z = 0.34$) was observed. The pairwise comparisons test for a significant three-way interaction [$F(1,1313) = 4.65$, $p = .031$] revealed that f_0 values were significantly higher for [i] responses in the condition of lower height (Height 2) when Location was above (M = 0.69 st) rather than below (M = 0.52 st) (p = .012; $d_z = 0.37$). This suggests that the f_0 values of [i] responses are

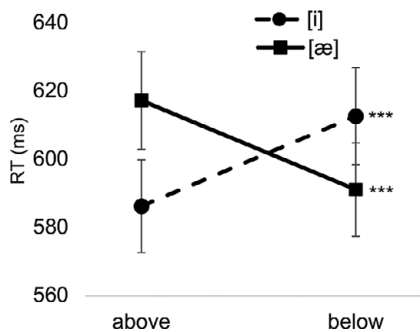

Figure 5. The mean vocal reaction times for Experiment 3 as a function of the response (vowel) and the above/below position of the target stimulus. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences (***$p < .001$).

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

heightened when the response is performed to the above condition rather than the below condition, but this effect is not robustly observed in both Height conditions. Regarding *F1*, the main effect of Response [$F(1,2588) = 58.03$, $p < .001$] showed that [æ] responses were higher (M = 529 Hz) than [i] responses (M = 303 Hz) ($d_z = 1.14$). In addition, the interaction between Location and Response was significant [$F(1,2588) = 28.55$, $p < .001$]. *F1* was higher for [i] responses (notice that the effect size is very small) when the location was above (M = 308 Hz) rather than below (M = 297) (p = .012; $d_z = 0.08$), while the effect was not significant for [æ] responses (p = .132). Regarding *F2*, the main effect of Response [$F(1,2674) = 1398.64$, $p < .001$] showed that [i] responses were higher (M = 2861 Hz) than [æ] responses (M = 1427 Hz) ($d_z = 5.7$). In addition, the interaction between Location and Response was significant [$F(1,2674) = 10.32$, $p < .001$]. *F2* values were higher for [æ] responses (notice that the effect size is very small) when the location was below (M = 2863 Hz) rather than above (M = 2858 Hz) (p < .001; $d_z = 0.02$), while the effect was not significant for [i] responses (p = .677).

Finally, 11 participants (47.8%) judged that the task was easier to perform in the [i]-above/[æ]-below block, and 5 participants (21.7%) judged that the task was easier to perform in the [i]-below/[æ]-above block. The rest of the participants did not report any difference between the blocks. Furthermore, two participants mentioned that they might have involuntarily raised/lowered their vocalization pitch according to the above/below location of the bird.

3.3. Discussion

The results of Experiment 3 replicated the congruency effect observed in the response times of Experiment 1. The [i] responses were produced faster when they were performed to the above position of the target, while the [æ] responses were produced faster when they were performed to the below position of the target. However, the effect observed in Experiments 1 and 2 in which the f_0 values were heightened when the target appeared at the upper location in comparison to the lower location was not robustly observed in Experiment 3, albeit there was a hint of this effect. The f_0 values of [i] responses were heightened when the responses were performed to the above condition rather than the below condition, but this effect was only observed in the lower height condition.

4. General discussion

The study investigated whether the spatial concepts of *up/above* and *down/below* would be sound-symbolically associated with a high and low vowel, respectively. In line with our hypothesis, the study revealed that the production of the high-front vowel [i] is associated with the concepts of *up* and *above*, while the production of the low-front vowel [æ] is associated with the concepts of *down* and *below*. This effect was observed in speeded-up vocal responses when the responses were performed in congruent ([i]-up; [æ]-down) rather than incongruent conditions. In order to contrast this finding with previous observations, research has shown that spatial concepts of distance (Johansson & Zlatev, 2013; Rabaglia et al., 2016; Tanz, 1971; Vainio, 2021), size (Knoeferle et al., 2017; Sapir, 1929; Thompson & Estes, 2011), and front/back (Vainio et al., 2023) are similarly sound-symbolically associated with

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

particular speech sounds. Hence, the present finding expands our understanding of spatial sound symbolism suggesting that *up* and *down* concepts of vertical space can be also sound-symbolically associated with particular vowels. Furthermore, the study shows that this phenomenon is not tight to concepts of vertical movement, but the non-movement-related concepts *above* and *below* are similarly associated with the high and low vowels, respectively. This proposes that this congruency effect is based on relatively robust and generalizable processes of semantic representation.

Experiment 2 tested whether the same congruency effect, which was observed between particular vowels and spatial concepts in Experiment 1, would be replicated with other speech sounds that also have the potential to provide the MARC effect. In Experiment 2, the marked ([æ]) and non-marked ([i]) vowels were replaced by the marked ([f]) and non-marked ([v]) consonants, while other aspects of Experiment 2 remained identical to that of Experiment 1. The congruency effect was not replicated in Experiment 2. The results of Experiment 2 supported the assumption that rather than being based on the MARC effect, the vowel-height congruency effect reflects sound-symbolic connections between the production of a particular vowel and the processing of these spatial concepts.

The theoretically relevant question is why the production of the high and low vowels is associated with spatial concepts *up/above* and *down/below*, respectively. Although the findings of this study cannot state anything conclusive about the mechanisms behind the height-vowel congruency effect observed in reaction times of the present study, we would like to discuss two possible explanations that might shed light on these mechanisms. It is important to emphasize that both of these explanations can provide the mechanistic basis for the effect, as it has been recognized that two or more mechanisms can provide a basis for a single sound symbolism phenomenon (Sidhu & Pexman, 2018). The first explanation emphasizes how the associative network of semantic memory represents sensory properties that naturally occur in the environment and also in relation to our own behavior. According to this sensory account, these spatial sound symbolism phenomena are a consequence of an implicit tendency to imitate environmental sounds by producing corresponding speech sounds. This account has been provided to explain spatial sound symbolism phenomena such as size, so that speech sounds with relatively high f_0 (e.g., [i]) are associated with small size because smaller animals tend to produce sounds with higher f_0 than larger animals (e.g., Knoeferle et al., 2017; Ohala, 1994). This perspective can be also adapted to explain the sound symbolism effect observed in the present study in relation to vertical space. As already presented in the Introduction, given that people have a tendency to link high/low tones with high/low vertical space (Pisanski et al., 2017; Pratt, 1930; Shintel et al., 2006; Spence, 2019), and high/low vowels have a relatively higher/lower IVP (Sapir, 1989; Whalen & Levitt, 1995), it is possible that this resemblance results in an implicit tendency to associate high/low vowels with high/low vertical space.

A second reason to assume that high/low vowels would be associated with the concepts of high/low is linked to the hypothesis that action representations play a key role in representing the meaning of many concepts in general and abstract concepts in particular (Binder & Desai, 2011; Dreyer & Pulvermüller, 2018; Pulvermüller, 2018). This action account emphasizes a tendency to implicitly imitate environmental properties by producing body movements (e.g., articulatory gestures) that mirror these properties (see Vainio, 2021 for a review). If we apply this perspective to explain spatial sound symbolism phenomena related to size, it can be stated that the low-back

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

vowels implicitly mirror the largeness of the referent's size by increasing the oral cavity and the high-front vowels mirror the smallness of the referent by decreasing the oral cavity (Ramachandran & Hubbard, 2001; Sapir, 1929; Vainio, 2021). Regarding the present finding, this hypothesis assumes that representing the meaning of spatial concepts such as *up* and *down* is partially grounded in action representations that move body parts upward/downward. Given that high/low vowels are produced by moving the larynx, tongue, and jaw upward/downward (Fant, 1960; Honda et al., 1999; Ladefoged, 1968), it is possible that semantic knowledge of these spatial concepts is implicitly connected to the articulatory movements of high and low vowels resulting in the sound-symbolic association between high/low vowels and high/low concepts.

As already mentioned, it has been previously observed that people tend to produce slightly higher-pitched speech when vocalizing the sentence "It is going up" when the visual stimulus moves up and lower-pitched speech when vocalizing the sentence "It is going down" when the visual stimulus moves down (Shintel et al., 2006). In the present study, this phenomenon was observed when the participants were required to produce a vowel [i] or [æ] (Experiment 1) and a syllable [fe] or [ve] (Experiment 2) depending on whether the visual stimulus moved upward or downward. The rising pitch was observed with the vowel [i] and syllables [fe] and [ve] when the stimulus moved upward in comparison to downward. However, a similar effect was not observed in relation to the vowel [æ] suggesting that vowel articulation processes might have a role in this effect. The effect might be diminished with those vowel articulations that require down-directed movement of the larynx, tongue, and jaw from a speech-ready position (such as [æ]) that was adapted to the position optimally ready to produce both of the speech sounds in the experiment (cf. Simko & Cummins, 2010). If this interpretation of the results is correct, one might state that this vocal pitch-elevation effect is, at least to some extent, caused by the elevated movement of the vocal apparatuses that mirrors the elevated movement of the visual stimulus and which results in pitch raise of vocalization. Furthermore, Experiment 3 revealed that when these [i]/[æ] responses were performed according to the spatial concepts *above* and *below* there was only a hint of this effect. This might suggest that perhaps this effect is emphasized when speech is referring to absolute verticality (e.g., a bird moves up) rather than relative verticality (e.g., a bird is above a cloud). Finally, given that the participants mostly did not report being aware that they produced higher-pitched vocalizations when the stimulus moved upward suggests that this vocal pitch-elevation effect operates implicitly.

Nevertheless, similarly to the height-vowel congruency effect observed in the reaction times of the present study, the vocal pitch-elevation effect can be also explained by two possible mechanisms: the sensory account and the action account. The sensory account highlights a tendency to imitate environmental sounds by producing corresponding speech sounds resulting in uttering higher-pitched sounds in relation to the up-directed stimulus. This occurs perhaps because higher frequencies originate more frequently from elevated sources in the natural environment (Parise et al., 2014). The action account assumes that higher-pitched vocalizations are associated with higher vertical space, not solely because people tend to imitate environmental sounds by producing corresponding speech sounds per se, but rather because the relative raising or lowering of vocalization pitch is a consequence of vertically moving head, larynx, tongue, and jaw. Indeed, it has been shown that head elevation correlates with *f0* rise and head lowering with *f0* fall (Cwiek & Fuchs, 2019;

https://doi.org/10.1017/langcog.2023.31 Published online by Cambridge University Press

Liu et al., 2020; Munhall et al., 2004) perhaps because head elevation changes the position of the larynx, which in turn pulls on the cricothyroid muscle leading to f_0 rise (Honda et al., 1999). Correspondingly, the intentional production of a high-pitched voice might involve lifting the larynx and the tongue (Higashikawa et al., 1996; Saldías et al., 2021), and high vowels, which typically have a relatively high intrinsic pitch, are produced by moving the larynx, tongue, and jaw upward (Fant, 1960; Honda et al., 1999; Ladefoged, 1968). In general, according to this account, high tones have the connotation of 'being high' because high tones are implicitly associated with up-directed movements of the head, tongue, jaw, and larynx. However, it is possible that both of these phenomena – acoustically and gesturally based – might contribute to the pitch-elevation phenomenon.

In conclusion, the study presents a novel sound-meaning correspondence effect in which the concepts of *up/above* and *down/below* are sound-symbolically associated with a high and low vowel, respectively. Given that sound-symbolic sound-meaning pairing has been also linked to the spatial concepts of size (Knoeferle et al., 2017; Sapir, 1929; Thompson & Estes, 2011), distance (Johansson & Zlatev, 2013; Rabaglia et al., 2016; Tanz, 1971; Vainio, 2021), and front/back (Vainio et al., 2023), it seems that spatial concepts are particularly prone to be represented in an iconic manner in speech. Although this study did not reveal the mechanisms behind the effect, we proposed that the acoustic and articulatory aspects of high and low vowels can contribute to this effect. Furthermore, the study replicated the vocal pitch-elevation effect (Shintel et al., 2006) so that people tended to utter higher-pitched vocalizations in relation to the up-directed stimulus. This observation shows that the pitch-elevation effect, observed in a vocal context, is a robust phenomenon.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/langcog.2023.31.

Data availability statement. The data is available at https://osf.io/wkqrf/?view_only=2deb7f3c3d814050a7ee9b860ed6085f.

Acknowledgments. We would like to thank Dr. Markku Kilpeläinen for his help in analyzing the data.

References

Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, 175, 122–130.

Anikin, A., & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics*, 81(3), 764–777.

Bennett, D. A. (2001). How can I deal with missing data in my study?. *Australian and New Zealand Journal of Public Health*, 25(5), 464–469.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15 (11), 527–536.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9), 341–345.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 9.

Clark, N., Perlman, M., & Johansson Falck, M. (2013). Iconic pitch expresses vertical space. In *Language and the Creative Mind* (pp. 393–410). CSLI Publications.

Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5(1), 39–62.

Cwiek, A., & Fuchs, S. (2019). Iconic prosody is rooted in sensori-motor properties: Fundamental frequency and the vertical space. In *CogSci 2019: 41st annual meeting of the cognitive science society* (pp. 1572–1578).

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615.

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system?–An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, 100, 52–70.

Fant, G. (1960). *Acoustic theory of speech production*. Mouton.

Franca, M., Turella, L., Canto, R., Brunelli, N., Allione, L., Andreasi, N. G., Desantis, M., Marzoli, D., & Fadiga, L. (2012). Corticospinal facilitation during observation of graspable objects: A transcranial magnetic stimulation study. *PLoS One*, 7(11), e49025.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Higashikawa, M., Nakai, K., Sakakura, A., & Takahashi, H. (1996). Perceived pitch of whispered vowels-relationship with formant frequencies: A preliminary study. *Journal of Voice*, 10(2), 155–158.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, 42(4), 401–411.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130298.

Johansson, N., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition*, 12(1), 56–83. https://doi.org/10.1017/langcog.2019.35

Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *The Public Journal of Semiotics*, 5(1), 1–20.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, 7(1), 1–11.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--A model and taxonomy. *Psychological Review*, 97(2), 253.

Körner, A., & Rummer, R. (2022). Articulation contributes to valence sound symbolism. *Journal of Experimental Psychology: General*, 151(5), 1107.

Ladefoged, P. (1968). *A phonetic study of West African languages: An auditory-instrumental survey* (Vol. 1). Cambridge University Press.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 726.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.

Liu, Y., Shamei, A., Chow, U. Y., Soo, R., Pineda Mora, G., de Boer, G., & Gick, B. (2020). F0-related head movement in blind versus sighted speakers. *The Journal of the Acoustical Society of America*, 148(2), EL190–EL194.

Margiotoudi, K., & Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports*, 10(1), 1–13.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1152.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15(2), 133–137.

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, 4 (2), 115–125.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound Symbolism*, 2, 325–347.

Pääkkönen, M. (1991). A:sta Ö:hön. Suomen kielen yleisyystilastoja. *Kielikello*, 1.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(16), 6104–6108.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1, 227.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130300.

Pisanski, K., Isenstein, S. G., Montano, K. J., O'Connor, J. J., & Feinberg, D. R. (2017). Low is large: Spatial location and pitch interact in voice-based body size estimation. *Attention, Perception, & Psychophysics*, 79, 1239–1251.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, 13(3), 278–285.

Pulvermüller, F. (2018). The case of CAUSE: neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170129.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865–7870.

Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., & Trope, Y. (2016). The sound of distance. *Cognition*, 152, 141–149.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--A window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3–34.

Saldías, M., Laukkanen, A. M., Guzmán, M., Miranda, G., Stoney, J., Alku, P., & Sundberg, J. (2021). The vocal tract in loud twang-like singing while producing high and low pitches. *Journal of Voice*, 35(5), 807.e1–807.e23.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3), 225. https://doi.org/10.1037/h0070931

Sapir, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. *Journal of Voice*, 3(1), 44–51.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55(2), 167–177.

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, 25, 1619–1643.

Simko, J., & Cummins, F. (2010). Embodied task dynamics. *Psychological Review*, 117(4), 1229.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus- response compatibility: An integrated perspective. Advances in Psychology* (Vol. 65, pp. 31–86). Amsterdam.

Spence, C. (2019). On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory Research*, 32(3), 235–265.

Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, 14(3), 266–276.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, 64(12), 2392–2404.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual cognition*, 8(6), 769–800.

Vainio, L. (2021). Magnitude sound symbolism influences vowel production. *Journal of Memory and Language*, 118, 104213.

Vainio, L., Kilpeläinen, M., Wikström, A., & Vainio, M. (2023). Sound-space symbolism: Associating articulatory front and back positions of the tongue with the spatial concepts of forward/front and backward/back. *Journal of Memory and Language*, 130, 104414.

Vainio, L., Rantala, A., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2017). Systematic influence of perceived grasp shape on speech production. *PloS One*, 12(1), e0170221.

Vainio, L., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2015). Interaction in planning movement direction for articulatory gestures and manual actions. *Experimental Brain Research*, 233, 2951–2959.

Vainio, L., Tiainen, M., Tiippana, K., & Vainio, M. (2019). Connecting directional limb movements to vowel fronting and backing. *Neuroscience Letters*, 711, 134457.

Vainio, L., Tiippana, K., Tiainen, M., Rantala, A., & Vainio, M. (2018). Reaching and grasping with the tongue: Shared motor planning between hand actions and articulatory gestures. *Quarterly Journal of Experimental Psychology*, 71(10), 2129–2141.

Vainio, L., & Vainio, M. (2021). Sound-action symbolism. *Frontiers in Psychology*, 12, 718700.

Welford, A. T. (1980). *Reaction times*. Academic Press.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Willmes, K., & Iversen, W. (1995). On the internal representation of number parity. In *Paper presented at the spring annual general meeting of the British neuropsychological society, London*.

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, 57, 209–219.

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, 6(1), 1–13.

Zimmer, K. (1964). Affixed negation in English and other languages: An investigation of restricted productivity. Word, 20, 2, Monograph No. 5.

Cite this article: Vainio, L., Wikström, A., Repetto, C., & Vainio, M. (2023). Sound-symbolic association between speech sound and spatial meaning in relation to the concepts of *up/down* and *above/below*, *Language and Cognition* 15: 884–903. https://doi.org/10.1017/langcog.2023.31