The Bouba–Kiki effect is predicted by sound properties but not speech properties

Ananya Passi[1] · S. P. Arun[2]

Abstract

Humans robustly associate spiky shapes to words like "Kiki" and round shapes to words like "Bouba." According to a popular explanation, this is because the mouth assumes an angular shape while speaking "Kiki" and a rounded shape for "Bouba." Alternatively, this effect could reflect more general associations between shape and sound that are not specific to mouth shape or articulatory properties of speech. These possibilities can be distinguished using unpronounceable sounds: The mouth-shape hypothesis predicts no Bouba-Kiki effect for these sounds, whereas the generic shape-sound hypothesis predicts a systematic effect. Here, we show that the Bouba–Kiki effect is present for a variety of unpronounceable sounds ranging from reversed words and real object sounds ($n = 45$ participants) and even pure tones ($n = 28$). The effect was strongly correlated with the mean frequency of a sound across both spoken and reversed words. The effect was not systematically predicted by subjective ratings of pronounceability or with mouth aspect ratios measured from video. Thus, the Bouba–Kiki effect is explained using simple shape-sound associations rather than using speech properties.

Introduction

Languages often contain systematic associations between object names and their visual properties (Dingemanse et al., 2015; Lockwood & Dingemanse, 2015; Sidhu & Pexman, 2018). Indeed, there are many cross-modal associations between shapes and sounds (Spence, 2011). A famous example is the Bouba–Kiki effect, whereby people associate rounded shapes to words like "Bouba" and spiky shapes to words like "Kiki" (Fig. 1a). This effect, first reported by Köhler using the words "baluba" and "takete" (Köhler, 1967), was subsequently termed the Bouba–Kiki effect (Ramachandran & Hubbard, 2001). This effect has since been reported robustly across diverse populations (Bremner et al., 2013; Chen et al., 2019; Davis, 1961; Hung et al., 2017; Sučević et al., 2015) with a few exceptions (Gold & Segal, 2017; Oberman & Ramachandran, 2008; Occelli et al., 2013; Rogers & Ross, 1975; Styles & Gawne, 2017). It has also been observed in preverbal children and infants (Asano et al., 2015; Imai et al., 2015; Maurer et al., 2006; Ozturk et al., 2013). This effect is predicted by specific articulatory features (stop consonants or close front vowels for Kiki-like words, continuant consonants or open back vowels for Bouba-like words) in sounds (D'Onofrio, 2014; Fort et al., 2015; Knoeferle et al., 2017; Westbury et al., 2018) and specific visual features (low spatial frequency for Bouba-like and high spatial frequency for Kiki-like) in shapes (Chen et al., 2021, Chen et al., 2016; Kim, 2020; Turoman & Styles, 2017).

How does the Bouba–Kiki effect arise? A popular account suggests that the Bouba–Kiki effect arises due to neurological cross-activation between motor and auditory
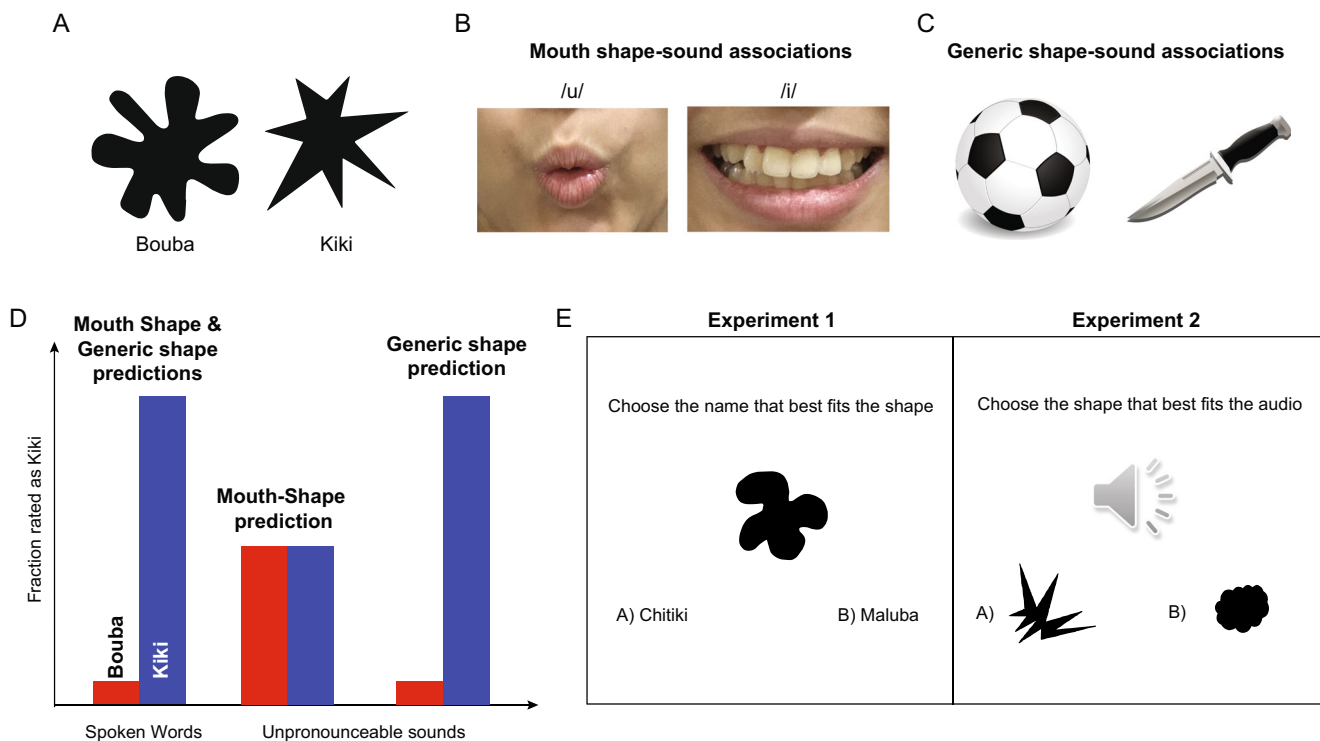
✉ S. P. Arun
sparun@iisc.ac.in

[1] Undergraduate Programme, Indian Institute of Science, Bengaluru, India

[2] Centre for Neuroscience, Indian Institute of Science, Bengaluru, India

🙋 Springer

Fig. 1 The Bouba–Kiki effect and its explanations. **a** Across cultures and even at a young age, most people will associate rounded shapes to words like "Bouba" and spiky shapes to words like "Kiki." Image source: Ramachandran and Hubbard (2001). **b** According to the mouth-shape explanation, the mouth forms a rounded shape while uttering "Bouba" and an angular shape while uttering "Kiki," and the Bouba–Kiki effect reflects these associations. Image source: Authors own photograph. **c** According to the generic shape explanation, real-world objects make systematically different sounds depending on their shapes, and the Bouba–Kiki effect reflects these associations. Image source: Wikimedia Commons. **d** The mouth shape and generic shape hypotheses both explain the Bouba–Kiki effect observed for spoken words (*left bars*) but yield different predictions for unpronounceable sounds such as reversed words, real object sounds, and pure tones (*middle and right bars*). Since such sounds cannot be pronounced, the mouth-shape hypothesis predicts no effect (*middle bars*). By contrast, since such sounds can be produced by real-world objects, the generic shape hypothesis predicts systematic associations (*right bars*). **e** Experimental design for key experiments: In Experiment 1 (*left panel*), we validated the Bouba–Kiki effect by presenting a shape along with two words (one Bouba-like and one Kiki-like word) and asking participants to choose the word that best fits the shape. In Experiments 2 & 3 (*right panel*), we played a sound and presented two shapes (one round, one spiky shape), and participants had to indicate the shape best fits the sound. (Color figure online)

cortices. In other words, representations of lip/tongue movements may be nonarbitrarily mapped to certain sounds. Since our mouths make rounded shapes on uttering Bouba-like words and angular shapes on saying Kiki-like words (Fig. 1b), we learn to associate these shapes to these words (Ramachandran & Hubbard, 2001; Sapir, 1929). Evidence in favour of this account comes from associations between happy/sad emotions and Kiki/Bouba-like words regardless of whether the emotion is conveyed through round or angular features (Karthikeyan et al., 2016; Sievers et al., 2019). However, this observation could reflect higher level associations between facial emotion and sounds. An alternate possibility is that the Bouba–Kiki effect reflects generic associations between the shapes of objects and the sounds they produce that are not specific to speech (Fig. 1c). This possibility is supported by the presence of the Bouba–Kiki effect in preverbal infants (Asano et al., 2015; Imai et al., 2015; Maurer et al., 2006; Ozturk et al., 2013; P. Walker et al.,

2010), and also by the many other general cross-modal correspondences reported between vision and sound (Albertazzi et al., 2015; Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Cowles, 1935; Evans & Treisman, 2010; Gallace & Spence, 2006; Guzman-Martinez et al., 2012; Hubbard, 1996; Ković et al., 2017; Liew et al., 2018; Lim & Styles, 2016; Ludwig et al., 2011; Marks, 1987; O'Boyle & Tarte, 1980; Parise & Spence, 2012; Walker et al., 2012).

To summarize, the Bouba–Kiki effect can be explained either as a specific association between mouth shape and sound, or as a more general association between object shape and sound. A critical test of these two explanations is sounds that are hard to pronounce such as reversed words or environmental sounds (Fig. 1d). According to the mouth-shape hypothesis, since such sounds cannot be articulated easily, the Bouba–Kiki effect should be abolished or reduced in magnitude. By contrast, the generic shape-sound hypothesis predicts a robust Bouba–Kiki effect even for such sounds.

Overview of this study

In this study, we set out to discriminate between these two hypotheses by measuring the Bouba–Kiki effect on unpronounceable sounds. In Experiment 1, we confirmed the Bouba–Kiki effect to be present on a novel set of 20 rounded/spiky shapes and 20 Bouba–Kiki words. In Experiment 2, we measured the Bouba–Kiki effect on reversed words created by playing each spoken word backwards in time. These sounds have identical frequency content as the original words, while at the same time being less pronounceable, making them ideal probes for testing the mouth-shape hypothesis. We also measured the Bouba–Kiki effect on high/low pitched natural sounds recorded from real-world objects when they are struck. Since these sounds are also hard to pronounce, the mouth-shape hypothesis predicts that the Bouba–Kiki effect would be abolished.

In Experiment 3, we collected subjective ratings of pronounceability for spoken and reversed words to confirm that the reversed words were indeed harder to pronounce, and to investigate whether the Bouba–Kiki effect was predicted using pronounceability. In Experiment 4, we measured mouth shape from a participant's video and asked whether the Bouba–Kiki effect can be predicted using mouth aspect ratio. In Experiment 5, we measured the Bouba–Kiki effect for pure tones varying in frequency, to confirm that the effect was driven by sound properties rather than speech articulatory properties.

Experiment 1. Bouba-Kiki effect

Before testing the Bouba–Kiki effect on unpronounceable sounds, we set out to first confirm the effect to be present on the specific shapes and spoken words used in this study.

Methods

All participants gave informed consent to experimental protocols approved by the Institutional Human Ethics Committee at the Indian Institute of Science, Bangalore.

Participants A total of 45 participants were recruited through email advertisements in our institute mailing lists (22 female, 21 male, 2 undisclosed; age = 18–78 years; mean ± SD: 30 ± 19 years). We selected this sample size because previous studies of the Bouba–Kiki effect have yielded clear effect sizes with similar or even smaller numbers of participants (Bremner et al., 2013; Kim, 2020). Moreover, we observed consistent results across subjects in our experiment (see Results). This confirmed that the selected sample size was sufficient for the given experiment.

Since language experience could potentially influence the results, we performed a post hoc assessment to report the linguistic background of the participants through a survey on Google Forms. Each participants were asked to report the languages they were familiar with and their proficiency in each language on a scale of 1 to 10 (1 = *poor*, 10 = *highly fluent*). A majority of the participants (60%, $n = 27$ of 45) completed the survey. All participants were highly fluent in English (self-reported fluency, mean ± SD: 8.5 ± 1.2) and were additionally fluent in a number of other languages (median number of languages = 3, self-reported fluency in first, second and third languages: mean ± SD: 8.4 ± 1.2 for first language across 27 participants, 6.5 ± 2.1 for second language across 25 participants, 5.0 ± 2.0 for third language across 17 participants. Hindi was the most common language in which participants reported being the most fluent ($n = 19$ participants), followed by Kannada ($n = 4$).

Stimuli We created 20 black shapes on a white background using Microsoft Paint. Each stimulus was created with a resolution of 650 × 601 pixels. Of these, 10 had rounded protrusions and 10 had spiky protrusions (Fig. 2a). Similarly, we created 20 pseudowords, of which 10 were designed to be "Kiki"-like and 10 were designed to be "Bouba"-like (Fig. 2a) using vowels/consonants previously reported as being so. Specifically, Bouba-like words contained vowels like /u/, /o/ and consonants like /m/, b/ while Kiki-like words contained vowels like /i/, /e/ and consonants like /k/, /t/.

Procedure We conducted this experiment using Google Forms. On each trial, depicted schematically in Fig. 1e, a new screen would appear, and participants saw one shape at a time, alongside two visually presented pseudowords as options (one Bouba-like word, one Kiki-like word). These word pairs were fixed across participants. We could not present spoken words as response options due to technical limitations of Google Forms. Trials appeared in random order for each participant. Participants were asked to select the pseudowords that best fit the shape. Each shape was presented exactly once, and the 10 "Kiki"-like and 10 "Bouba"-like pseudowords were seen twice during the entire experiment (10 × 2 = 20). Thus, the entire experiment consisted of only 20 trials, one for each shape. At the end of the experiment, 4.4% (2 of 45) participants reported that they had heard of the Bouba–Kiki effect before. We obtained qualitatively similar results upon excluding these participants.

In subsequent experiments, we realized that we had misclassified two words: *caucau*, a Kiki-like word, was classified in this experiment as being Bouba-like, whereas *bemele*, a Bouba-like word, was classified as being Kiki-like. Due to this misclassification, the experiment consisted of two trials
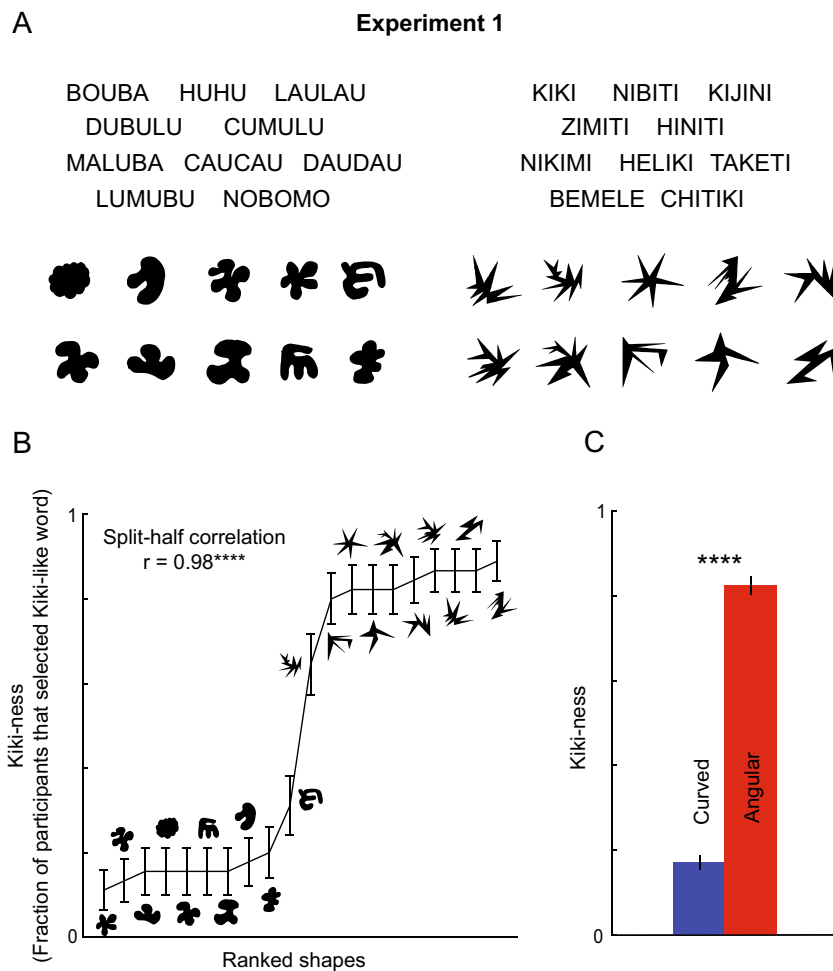
Fig. 2 Bouba–Kiki effect validation (Experiment 1). **a** Pseudowords and shapes used in Experiment 1. *Left column*: Bouba-like words and shapes. *Right column*: Kiki-like words and shapes. Pronunciation transcripts for Bouba-like words: /buːbɑ/, /hʊhʊ/, /laʊlaʊ/, /duːbuːluː/, /kʊmʊlʊ/, /məlʊbaː/, /kaʊkaʊ/, /daʊdaʊ/, /lʊmʊbʊ/, /noʊboʊmoʊ/. Pronunciation transcripts for Kiki-like words: /kɪkiː/, /nɪbɪtiː/, /kɪdɪ ni:/, /zɪmɪti:/, /hɪnɪti:/, /nɪkɪmi:/, /heɪlɪki:/, /tɑːkeə ti:/, /beɪmeɪleɪ/, /tʃɪtɪki:/. **b** Kikiness (fraction of participants who selected a Kiki-like word) plotted for each shape. Error bars represent standard error of mean across participants calculated after scoring responses as 0 for Bouba and 1 for Kiki. The correlation coefficient between the Kikiness from odd- and even-numbered participants is shown (*top left*). Asterisks beside the correlation coefficient indicates its statistical significance (**** is $p < .0005$). **c** Average Kikiness for curved (blue) and angular (red) shapes. Error bars represent standard error of mean across shapes of the Kikiness averaged across participants. Asterisks indicate statistical significance of the comparison (rank-sum test across Kikiness for 10 curved and 10 angular shapes, **** is $p < .0005$). (Color figure online)

with a pair of Bouba-like words and two trials with a pair of Kiki-like words as choices. Despite this, the rounded/angular shapes presented in these trials were classified as Bouba-/Kiki-like respectively. This was because *caucau* was more Bouba-like than the Kiki-like words that it was paired with. Similarly, *bemele* was more Kiki-like that the Bouba-like words it was paired with.

Data analyses Across participants, we analyzed the pseudoword options selected for each shape. To quantify the Bouba–Kiki association for a particular shape, we calculated the fraction of participants who selected a Kiki-like word for that shape, and this is denoted throughout as its "Kikiness."

Results

The shapes and words used in this experiment are shown in Fig. 2a. For each shape, we calculated the fraction of participants who selected a Kiki-like word and denoted this as the Kikiness. The Kikiness for shapes sorted in ascending order is shown in Fig. 2b, revealing a clear difference between the responses for the rounded and angular shapes. To confirm the reliability of this measure across subjects, we calculated the Pearson's correlation in the Kikiness obtained from odd- and even-numbered participants. This revealed a high and statistically significant correlation across the entire set of shapes ($r = .98$, $p < .0005$). This suggested that all participants had a high degree of agreement with each other. This consistency also validates this measure, Kikiness, to be a reasonable

measure of the Bouba–Kiki effect strength. A large Kikiness for a given shape therefore indicates that nearly all participants selected a Kiki-like word.

Having established that participants responses were consistent, we proceeded to ask whether the responses confirmed the presence of the Bouba–Kiki effect on our stimuli. As expected, the Kikiness was significantly higher for the angular shapes compared with the rounded shapes (Fig. 2c; mean ± SD: 0.17 ± 0.05 for rounded shapes, 0.82 ± 0.07 for angular shapes, $p < .0005$, rank-sum test on mean Kikiness for the 10 rounded and 10 angular shapes).

The above shape-sound associations could be driven by angularity of shapes or by their size. To assess this possibility, we calculated the total area of each shape and asked whether this was correlated with the observed Kikiness of these shapes. This revealed an overall negative correlation ($r = −.89$, $p < .0005$), because the Bouba-like shapes generally occupied more area than the Kiki-like shapes. However we observed no systematic within-category correlations for Bouba-like or Kiki-like shapes considered separately (correlation between Kikiness and area: $r = .43$, $p = .21$ for Bouba-like shapes; $r = −.27$, $p = .45$ for Kiki-like shapes).

Conclusions

We conclude that participants systematically associated rounded shapes to Bouba-like words and angular shapes to Kiki-like words, confirming the Bouba–Kiki effect for these particular shapes and words.

Experiment 2. Reversed words and real object sounds

The results of Experiment 1 demonstrate that the Bouba–Kiki effect is indeed present for the specific shapes and words in our experimental paradigm. In Experiment 2, we sought to distinguish between the mouth-shape and generic-shape hypotheses by measuring the Bouba–Kiki effect for unpronounceable sounds. We tested three types of sounds: spoken words (those validated in Experiment 1), reversed versions of these words, and real object sounds. The real object sounds comprised low- and high-frequency sounds made by real-world objects when they were struck (e.g., pillow vs. metal objects).

Methods

All procedures were similar to Experiment 1, except for those detailed below.

Participants A total of 45 participants were recruited for this experiment (21 female, 24 male; age = 18–73 years; mean ± SD: 28.0 ± 15.3 years)—20 of these participants had also participated in Experiment 1. Excluding them yielded qualitatively similar results.

Since syllable experience could potentially influence the results, we performed a post hoc assessment of the linguistic background of the participants through a survey on Google Forms. Each participants were asked to report the languages they were familiar with and their proficiency in each language on a scale of 1 to 10 (1 = *poor*, 10 = *highly fluent*). A majority of the participants (75%, $n = 34$ of 45) completed the survey. All participants were highly fluent in English (self-reported fluency, mean ± SD: 8.5 ± 1.4) and were additionally fluent in a number of other languages (median number of languages = 3, self-reported fluency in first, second and third languages: mean ± SD: 8.5 ± 1.3 for first language across 34 participants, 6.5 ± 2.0 for second language across 31 participants, 4.9 ± 2.2 for third language across 18 participants). Hindi was the most common language in which participants reported being the most fluent ($n = 24$ participants), followed by Kannada ($n = 5$ participants).

Stimuli We created audio recordings of the 20 pseudowords used in Experiment 1 in a noise-free environment using built-in laptop microphones (voice of A.P.) at a sampling rate of 48000 Hz. To create the reversed words, we imported each audio recording into MATLAB, reversed the audio (*flip* function, MATLAB R2020a) and exported the audio into a file (*audiowrite* function, MATLAB R2020a). We selected 20 real object sounds from the audiovisual database The Greatest Hits (Owens et al., 2016). These included 10 high-frequency "metallic" sounds of various metal objects (e.g., kettles, railings) being struck with a drum stick, and 10 low-frequency "thud" sounds created by various upholstery items (e.g., sofas, cushions) being struck with the same drum stick. These objects were selected such that there was significant variation in their geometric structure and thus the sound they produced when struck. The shapes used were the same as in Experiment 1.

Procedure This experiment was conducted using Google Forms. On each trial, participants were given an audio recording along with two shapes (one rounded, one angular) as choices (Fig. 1e). Participants were asked to choose the shape that fits best with the audio. The audio recordings included 20 spoken words (10 Bouba-like and 10 Kiki-like words, same as those in Experiment 1—with the *caucau/bemele* misclassification resolved), 20 reversed words (reversed versions of the 10 Bouba-like and 10 Kiki-like words) and 20 real object sounds (10 low-frequency, 10 high-frequency). Trials were shown in random order, but the order was fixed across participants. Each audio recording was presented exactly once, and

the 10 angular and 10 rounded shapes were presented a total of 6 times. Thus the entire experiment involved 60 trials. At the end of the experiment, 15.6% (7 of 45) participants indicated that they had heard of the Bouba–Kiki effect before. We obtained qualitatively similar results upon excluding these participants.

Data analysis As before, we calculated Kikiness for each audio stimulus as the fraction of participants that selected the spiky shape. We calculated the mean frequency for each audio clip by taking its Fourier power spectrum, normalizing it so that it becomes like a probability distribution, and calculating the mean of this distribution. Using the frequency corresponding to the peak of the power spectrum did not yield a significant association with Kikiness.

Results

Participants were highly consistent in their responses as before (split-half correlation between Kikiness from odd- and even-numbered participants: $r = .89$ across all stimuli; $r = .88$ for spoken words; $r = .92$ for reversed words; $r = .90$ for real object sounds; $p < .0001$ in all cases).

In keeping with the classic Bouba–Kiki effect, the Kikiness for the Bouba-like words were significantly smaller than for Kiki-like words (Fig. 3a; mean $\pm$ *SD*: 0.15 ± 0.08 for Bouba-like words; 0.64 ± 0.09 for Kiki-like words; $p < .001$, rank-sum test comparing Kikiness for the two groups). This confirms the Bouba–Kiki effect for the spoken words.

Next we examined the Kikiness for the reversed words. Strikingly, here too, we observed a significantly larger Kikiness for the reversed Kiki-like words compared with the reversed Bouba-like words (Fig. 3a; mean $\pm$ *SD*: 0.18 ± 0.08 for reversed "Bouba"-like words, 0.56 ± 0.14 for reversed Kiki-like words, $p < .001$, rank-sum test). Even for real-object sounds, we observed a significantly larger Kikiness for metallic objects compared with cushion-like objects (Fig. 3a; mean $\pm$ *SD*: 0.48 ± 0.14 for cushion-like objects, 0.89 ± 0.06 for metallic objects, $p < .001$, rank-sum test).

Although the reversed Kiki-like words are associated with Kiki-like shapes easily, it is possible that their Kikiness changed in their quality compared with the original word. To assess this possibility, we plotted the Kikiness of the reversed words compared with the original spoken words. This revealed a strong significant correlation ($r = .86$, $p < .0005$; Fig. 3b), suggesting that participant responses captured audio features independent of playback direction.

The above results suggest that there could be simple audio features that determine the responses given by participants. As an initial step, we calculated the mean frequency of each audio clip in the experiment (see Methods) and asked whether this mean frequency was correlated with Kikiness. Interestingly,

we observed a significant correlation for spoken and reversed words (Fig. 3c; $r = .78$ for spoken words, $r = .69$ for reversed words, $p < .001$ in both cases). We also observed a significant correlation for real object sounds (Fig. 3d; $r = .89$, $p < .0001$). However, the range of mean frequencies for the real object sounds was larger compared with the spoken words, suggesting that mean frequency alone cannot fully account for the Kikiness and that perhaps other spectral properties might be required to do so.

Conclusions

We conclude that the Bouba–Kiki effect is present for sounds with varying degrees of pronounceability. This suggests that pronounceability of a word is not a prerequisite for the Bouba–Kiki effect, contradicting the mouth-shape hypothesis. We also show that spectral properties of sounds (such as their mean frequencies) can be reliable predictors of the effect strength, consistent with the generic shape-sound hypothesis.

Experiment 3. Pronounceability

In the previous experiment, we found that the Bouba–Kiki effect is present for both spoken as well as reversed words. According to the mouth-shape hypothesis, differences in the ease of pronunciation could explain the Bouba–Kiki effect strength. Words that are easy to pronounce should show a stronger effect. To investigate this possibility, we collected subjective ratings from independent sets of human participants regarding the ease of pronunciation of both spoken and reversed words.

Methods

All procedures were similar to Experiment 2, except for those detailed below.

Participants Since spoken words and reversed words vary widely in their pronounceability, we collected subjective ratings from two separate sets of participants for these two word groups. For assessing spoken words, we recruited 22 participants (seven female, 15 male; age = 18–48 years; mean $\pm$ *SD*: 21.1 ± 6.3 years)—one of these participants had participated in Experiments 1 and 2. For assessing reversed words, we recruited 26 participants (10 female, 14 male, two undisclosed; age = 18–52 years; mean $\pm$ *SD*: 29.9 ± 10.8 years)—two of these participants had participated in Experiment 1 and two participants had participated in Experiment 2. We obtained qualitatively similar results on excluding these repeated participants.

Since syllable experience could potentially influence the results, we performed a post hoc assessment of the
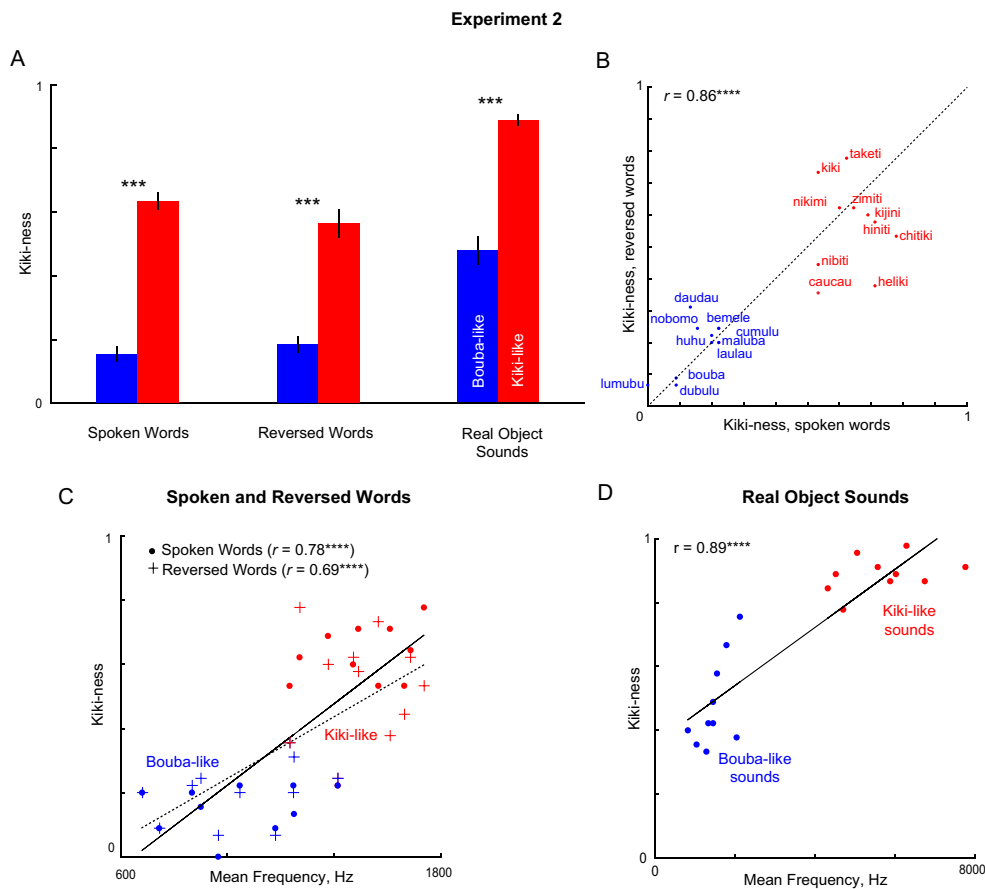
Fig. 3 Reversed words and real object sounds (Experiment 2). **a** Average Kikiness for Bouba-like (*blue*) and Kiki-like (*red*) sounds, for spoken words (*left bars*), reversed words (*middle bars*) and real object sounds (*right bars*). Asterisks represent statistical significance of each comparison (*** is $p < .001$). **b** Kikiness for each reversed word plotted against the Kikiness for the original spoken word. Asterisks beside the correlation coefficient indicate its statistical significance (**** is $p < .0005$). **c** Kikiness for spoken and reversed words against their mean frequency. The solid line indicates the best-fit line for spoken words, and the dashed line for reversed words. Asterisks are indicated as before. **d** Same as **c**, but for real object sounds. (Color figure online)

linguistic background of the participants through a survey on Google Forms. Each participants were asked to report the languages they were familiar with and their proficiency in each language on a scale of 1 to 10 (1 = *poor*, 10 = *highly fluent*). A majority of the participants (62.5%, *n* = 30 of 48) completed the survey. All participants were highly fluent in English (self-reported fluency, mean ± *SD*: 8.7 ± 1.1) and were additionally fluent in a number of other languages (median number of languages = 2.5, self-reported fluency in first, second and third languages: mean ± *SD*: 8.5 ± 1.6 for first language across 30 participants, 6.7 ± 2.1 for second language across 26 participants, 4.3 ± 1.6 for third language across 12 participants). Hindi was the most common language in which participants reported being the most fluent (*n* = 11 participants), followed by Kannada (*n* = 4 participants).

Stimuli We used the same audio recordings of 20 spoken and 20 reversed words as in Experiment 2.

Procedure Both tasks were conducted using Google Forms. An audio recording was played in each trial and participants were asked to rate its pronounceability on a scale of 1 (*easy to pronounce*) to 10 (*difficult to pronounce*). Trials were presented in a random order, but the order was fixed across participants. Each audio recording was presented exactly once. Thus, the entire experiment involved 20 trials (10 Kiki-like and 10 Bouba-like). At the end of the experiment, 27.3% (6 of 22) participants of the spoken word task and 26.9% (7 of 26) participants of the reversed word task indicated that they had heard of the Bouba–Kiki effect before. We obtained qualitatively similar results upon excluding these participants.

Data analysis We converted the subjective rating provided into a measure of ease of pronunciation by calculating 10 minus the rating provided by each participant. This measure, which we denote as "pronounceability rating" throughout, is large when the sound is easy to pronounce and small when it is hard to pronounce.

Results

Participants were highly consistent in their pronounceability ratings (correlation between mean ratings of odd- and even-numbered participants: $r = .85$ for spoken words; $r = .89$ for reversed words; $p < .0001$ in both cases). As expected, words became significantly less pronounceable when reversed (Fig. 4a; pronounceability ratings, mean $\pm$ SD: 7.2 ± 0.8 for spoken words, 4.0 ± 1.0 for reversed words, $p < .00005$, sign-rank test on mean ratings across all 20 words). This was true for Bouba-like and Kiki-like words considered separately as well (Fig. 4a). Interestingly, Bouba-like reversed words were rated as being more pronounceable compared with Kiki-like reversed words (pronounceability ratings, mean $\pm$ SD: 4.7 ± 0.5 for Bouba-like & 3.2 ± 0.7 Kiki-like words; $p < .0005$, unpaired t test). However no such difference in pronounceability was observed for spoken Bouba-like and Kiki-like words (pronounceability ratings, mean $\pm$ SD: 7.5 ± 0.6 for Bouba-like & 7.0 ± 0.8 Kiki-like words; $p = .15$, unpaired t test). This Bouba–Kiki difference in pronounceability was significantly different between spoken and reversed words (average difference in pronounceability, Bouba–Kiki, mean $\pm$ SD: 0.5 ± 0.32 for spoken words; 1.5 ± 0.26 for reversed words; both calculated by randomly sampling Bouba and Kiki words with replacement, and repeatedly calculating the mean pronounceability difference for 100,000 times; fraction of pairs in which this trend was reversed: $p = .007$).

The above results show that Kiki-like reversed words were rated as less pronounceable than Bouba-like reversed words. This is expected since Kiki-like words contain voiceless stop consonants (such as /p/,/t/,/k/) and unrounded front vowels (such as /i/), which are strongly altered upon reversing. By contrast sonorant consonants (e.g., /l/, /m/) and bilabial back vowels (e.g., /u/) are relatively less affected by reversing. We also observed a significant correlation between the pronounceability ratings for spoken and reversed words ($r = .55$, $p = .01$)—suggesting that words that are easy to pronounce are also easy to pronounce when reversed.

The above result can be explained by noting that some sounds are unaltered upon reversing, such as pure tones. Bouba-like words often contain sonorant consonants (e.g., /l/,/m/) and bilabial back vowels (e.g., /u/) which are closer to pure tones and therefore unaltered upon reversing. By contrast, Kiki-like words contain voiceless stop consonants (such as /p//t/,/k/) and unrounded front vowels (such as /i/)—which are drastically altered and become harder to pronounce upon reversing.

We also observed a significant correlation between the pronounceability ratings for spoken and reversed words ($r = .55$, $p = .01$)—suggesting that words that are easy to pronounce are also easy to pronounce when reversed. It is unclear to us why this might be so. We speculate that the factors that drive pronounceability are preserved upon reversal, such as perhaps the prolonged presence of certain sound frequencies.
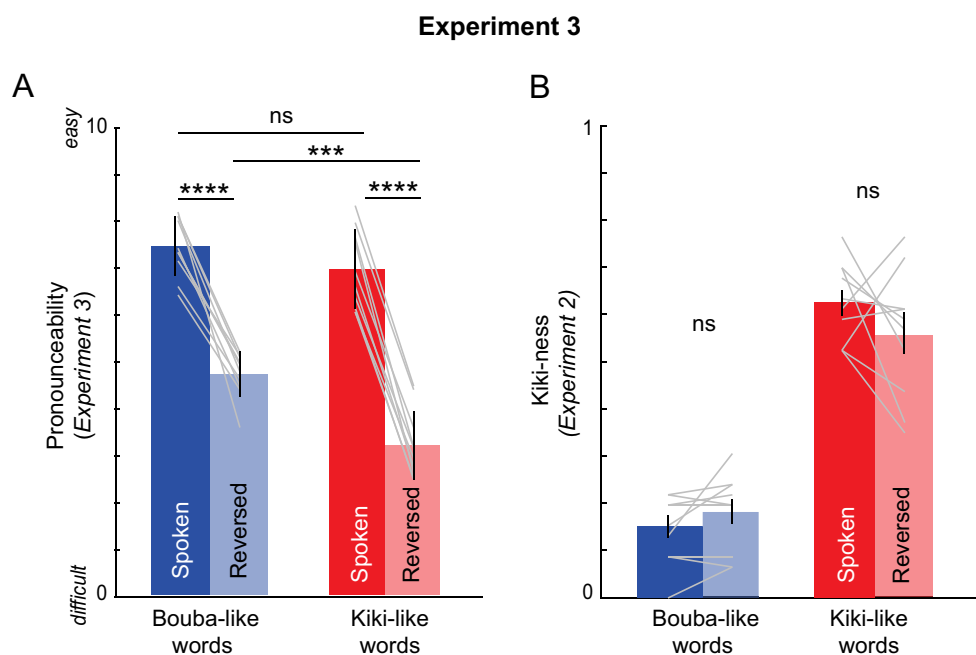
Experiment 3

Fig. 4 Subjective ratings of pronounceability (Experiment 3). **a** Average pronounceability for spoken (*dark*) and reversed (*light*) words, for Bouba-like (*blue*) and Kiki-like (*red*) words. *Light gray* lines indicate individual words and their reversed counterparts. Asterisks represent statistical significance of each comparison (* is $p < .05$, *** is $p < .0005$,

**** is $p < .00005$). **b** Average Kikiness for spoken (*dark*) and reversed (*light*) words, for Bouba-like words (*blue*) and Kiki-like (*red*) words. *Light gray* lines indicate individual words and their reversed counterparts. *ns*: not significant. (Color figure online)

Having characterized how subjective pronounceability ratings vary between spoken and reversed words and between Bouba-like and Kiki-like words, we next wondered whether these ratings are related to Kikiness observed in Experiment 2. In Experiment 2, we had found that the mean sound frequency was strongly correlated with Kikiness ($r = .74$, $p < .00005$ across both spoken and reversed words). Accordingly we wondered whether pronounceability directly predicts the Kikiness as well as mean frequency. Since Bouba-like words are considered more pronounceable than Kiki-like words whether they are spoken or reversed, we predicted a negative correlation between Kikiness and pronounceability. This was indeed the case: we observed a negative but not significant correlation between pronounceability and Kikiness for spoken words ($r = -.4$, $p = .07$), and a significant negative correlation for reversed words ($r = -.75$, $p < .0005$).

If pronounceability directly predicts the Bouba–Kiki effect, the drop in pronounceability from spoken to reversed words (Fig. 4a) should lead to a drop in the Kikiness from spoken to reversed words. To investigate this possibility, we plotted the Kikiness from Experiment 2 for Bouba-like and Kiki-like words separately (Fig. 4b). This revealed no significant difference in Kikiness between spoken and reversed Bouba-like words (Kikiness, mean ± SD: 0.15 ± 0.07 for spoken words, 0.18 ± 0.08 for reversed words, $p = .26$, sign-rank test across 10 words), or for Kiki-like words (Kikiness, mean ± SD: 0.64 ± 0.09 for spoken words, 0.56 ± 0.14 for reversed words, $p = .24$, sign-rank test). Likewise, at a more fine-grained level, the change in pronounceability for each word had no significant correlation with the change in Kikiness ($r = -.23$, $p = .15$).

Conclusions

Reversed words are less pronounceable than spoken words yet show an equally strong Bouba–Kiki effect. Thus, pronounceability cannot explain the Bouba–Kiki effect.

Experiment 4. Mouth shape

Here, we set out to further investigate the mouth-shape hypothesis by measuring the shape of the mouth from videos of a participant speaking these words, and asking whether fine-grained, word-level variations in the Bouba–Kiki effect can be explained using mouth shape.

Methods

Procedure A 34-year-old naïve female subject was recruited for this experiment. She was asked to serially read aloud a list of 20 pseudowords. This list consisted of the same 20 Kiki-like and Bouba-like words as in the previous experiments,

arranged in random order. Her lip movements were recorded using a mobile phone camera.

Data analysis Our video analysis is summarized in Fig. 5a. For each phoneme in each word, we captured a screenshot from the video-recording. From each screenshot, we measured the height and width of the mouth and calculated the mouth aspect ratio as the ratio of height to width. The aspect ratio for each word was taken as the average aspect ratios calculated across all constituent phonemes. We obtained qualitatively similar results upon taking the maximum instead of the average aspect ratio across phonemes.

Results

We calculated the aspect ratio of the mouth (height/width) for each word as the mean aspect ratio of mouth shape across the constituent phonemes, as depicted in Fig. 5a. Since the mouth forms a rounded shape while uttering phonemes associated with Bouba-like words, we predicted that the aspect ratio for Bouba-like words will be larger than for Kiki-like words. Indeed, the aspect ratio for Bouba-like words was significantly larger than the Kiki-like words (Fig. 5b; aspect ratio, mean ± SD: 0.69 ± 0.1 for Bouba-like words; 0.59 ± 0.1 for Kiki-like words; $p < .05$, t test across 10 words in each group).

Next we asked whether the aspect ratio predicts Kikiness for each word measured in Experiment 2. If mouth shape directly predicts Kikiness, then we would expect a robust negative correlation across words, since smaller aspect ratios correspond to Kiki-like words. However, we found no significant correlation between aspect ratio and Kikiness (Fig. 5c; $r = -.44$, $p = .051$ across all words. Moreover, upon considering Bouba-like and Kiki-like words separately, we observed no correlation in both cases (Fig. 5c; $r = .17$, $p = .64$ for Bouba-like words, $r = -.02$, $p = .95$ for Kiki-like words). These results are opposite to the prediction of the mouth-shape hypothesis at a fine-grained individual sound level.

Conclusions

Taken together, the above findings show that mouth shape, as measured using aspect ratio, does not systematically predict the Bouba–Kiki effect.

Additional analysis of Experiments 2–4

The results of Experiments 2–4 show that mean sound frequency predicts the strength of the Bouba–Kiki effect across individual words, but not pronounceability and mouth shape. These findings suggest that sound properties, rather than speech properties are sufficient to predict the Bouba–Kiki

Experiment 4

A

Aspect ratio calculation for maluba

/ma/ /lu/ /ba/

Aspect ratio calculation for taketi

/ta/ /ke/ /ti/

B

C
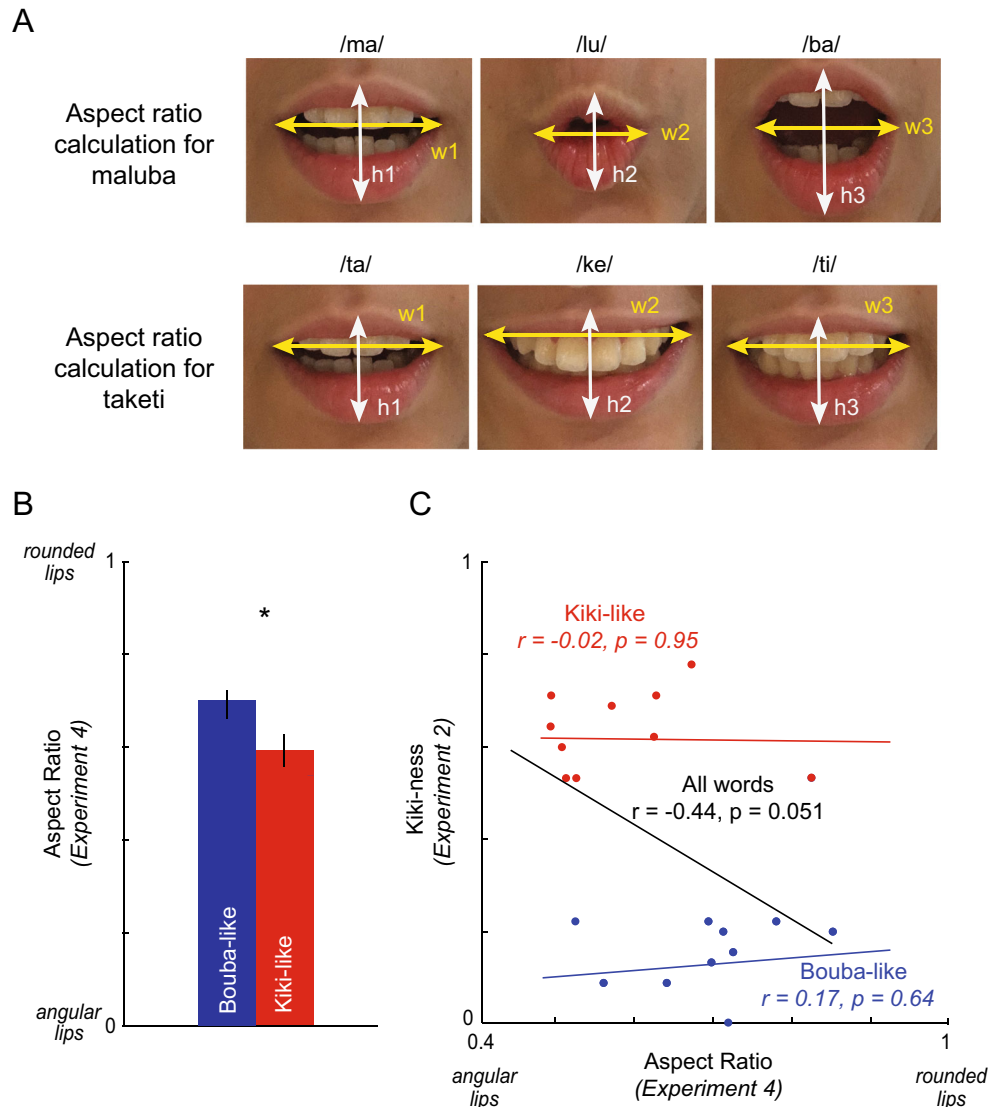
Fig. 5 Aspect Ratio (Experiment 4). **a** Schematic showing calculation of mouth aspect ratios for Kiki-like and Bouba-like words. Image source: A.P.'s own photograph. **b** Aspect Ratio for Bouba-like (*blue*) and Kiki-like (*red*) words. Asterisks represent statistical significance of the comparison (* is $p < .05$, *t* test). **c** Kikiness for Kiki-like and Bouba-like words (from Experiment 2) plotted against the mouth aspect ratio. The *black line* indicates the best-fit line for all words, the *blue line* for Bouba-like words and the *red line* for Kiki-like words. (Color figure online)

effect. However, it could be that some combination of pronounceability and mouth shape could still predict the effect. To investigate this possibility, we performed a computational analysis.

Results

We set out to investigate whether the strength of the Bouba–Kiki effect, as captured by the Kikiness, could be predicted using sound properties alone (mean frequency) as compared with speech properties (i.e., pronounceability and mouth aspect ratio).

To this end, for each of the 20 spoken words (10 Kiki-like + 10 Bouba-like) in Experiment 2, we asked whether the Kikiness of each word could be predicted using a weighted sum of the mean frequency, pronounceability and mouth aspect ratio. Specifically this meant fitting the Kikiness using a linear model of the form $y = Xb$ where y is a 20×1 vector containing the Kikiness of the 20 words, and X is a 20×4 matrix containing the mean frequency, pronounceability rating (from Experiment 3), mouth aspect ratio (Experiment 4) of each word in the first three columns and 1s along the last column. The 1 s in the last column represented the constant term in our linear regression. We solved this set of simultaneous equations using standard linear regression (*regress*

function in MATLAB) and calculated the predicted Kikiness. We repeated this analysis for five relevant combinations of models: (1) mean frequency alone; (2) pronounceability alone; (3) mouth shape alone; (4) speech properties (pronounceability + mouth shape); and (5) a combined model that included all these factors. Our goal was to compare models based on sound properties with those based on articulatory properties in terms of how well they predict Kikiness.

The resulting model fits are summarized in Fig. 6. The model based on mean frequency alone performed as well as the full model containing all factors, suggesting that the other factors (pronounceability and mouth shape) did not contribute further to the fit. Indeed, models based on speech properties (pronounceability, mouth aspect ratio or both) were all significantly poorer than the combined model that contained these factors together with mean frequency. Thus, sound properties strongly predict the Bouba–Kiki effect, and these predictions are not benefited by including speech-related properties like pronounceability and mouth aspect ratio. We obtained quantitatively similar results on performing a leave-one-out cross-validation.

Experiment 5. Pure tones

The results of Experiment 2 show that the Bouba–Kiki effect is present even for real object sounds. Here, we sought to
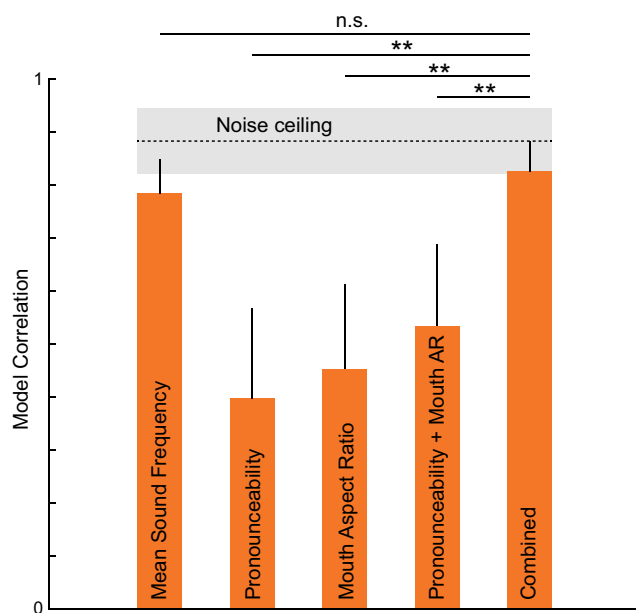
Fig. 6 Predicting the Bouba–Kiki effect using sound and speech properties. Model correlation between observed and predicted Kikiness are shown for models based on each factor. Error bars represent standard deviation, estimated by repeatedly sampling 20 words with replacement and repeating model fits for 10,000 times. Asterisks indicate statistical significance of comparing the full model fits on all 20 words with each individual model fit using a partial F test (** indicates $p < .005$, n.s. indicates $p > .05$). The noise ceiling is the split-half correlation in the subject responses

extend the generality of the findings in the preceding experiments to another class of unpronounceable sounds—namely, pure tones. To this end, we created pure tones with the same mean frequency as the spoken words in Experiments 1–2 and measured the Bouba–Kiki effect for these tones.

Methods

All procedures were identical to Experiment 2, and only the details specific to this experiment are summarized below.

Participants A total of 28 participants were recruited for this experiment (15 female, 13 male; age = 18–30 years; mean ± SD: 20.4 ± 2.3 years). None of these participants had also participated in Experiments 1 and 2.

Since syllable experience could potentially influence the results, we performed a post hoc assessment of the linguistic background of the participants through a survey on Google Forms. Each participants were asked to report the languages they were familiar with and their proficiency in each language on a scale of 1 to 10 (1 = *poor*, 10 = *highly fluent*). A majority of the participants (92.8%, n = 26 of 28) completed the survey. All participants were highly fluent in English (self-reported fluency, mean ± SD: 9.1 ± 1.2) and were additionally fluent in a number of other languages (median number of languages = 3, self-reported fluency in first, second and third languages: mean ± SD: 8.9 ± 1.2 for first language across 26 participants, 7.4 ± 1.8 for second language across 24 participants, 4.1 ± 1.5 for third language across 13 participants). Hindi was the most common language in which participants reported being the most fluent (n = 10 participants), followed by Bengali (n = 4).

Stimuli We created a set of 20 pure tones with a mean frequency equal to the spoken words used in Experiment 2 using the *audiowrite()* function from MATLAB 9.8 R2020a. These 20 tones along with the 20 spoken words formed the entire auditory stimulus set for this experiment. The spoken words were presented in this experiment to confirm that the participants exhibited the Bouba–Kiki effect.

Procedure Each of the forty audio clips were presented once in the course of the experiment with a unique pair of shapes as choices (one rounded, one angular). Each of the 10 angular and 10 rounded shapes were presented four times during the task. Thus the entire experiment consisted of 40 trials. At the end of the experiment, 21.4% (6 of 28) participants indicated that they had heard of the Bouba–Kiki effect before. We obtained qualitatively similar results upon excluding these participants.

Experiment 5
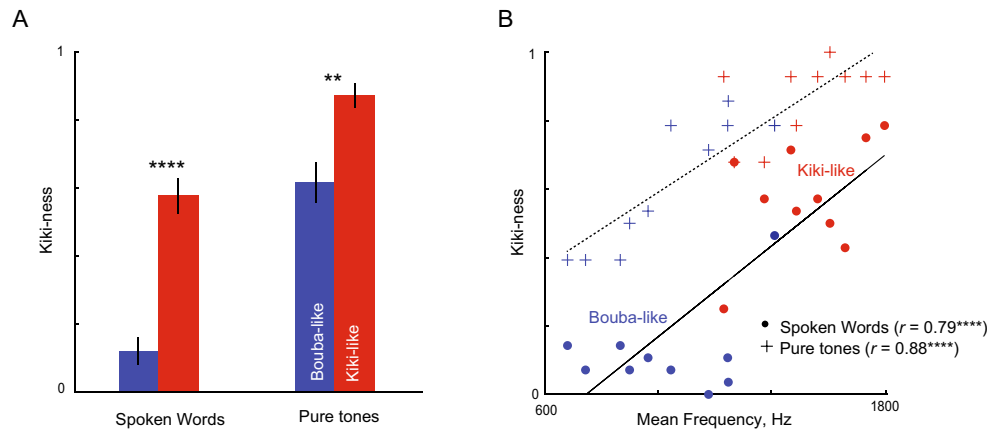
Fig. 7 Bouba–Kiki effect for pure tones (Experiment 5). **a** Kikiness for Bouba-like (*blue*) and Kiki-like (*red*) sounds for spoken words (*left bars*) and for pure tones (*right bars*). Asterisks represent statistical significance of each comparison (**** is $p < .0005$; ** is $p < .005$). **b** Kikiness for spoken words (*filled circles*) and pure tones (*crosses*) plotted against their mean frequency. Each pure tone was matched to have the same frequency as the corresponding Bouba-like or Kiki-like word. The *solid* and *dashed* line indicate the best linear fit for spoken words and pure tones respectively. (Color figure online)

Results

As before, participants were highly consistent in their responses (split-half correlation between Kikiness of odd- and even-numbered participants, $r = 0.93$ across all stimuli; $r = .92$ for spoken words; $r = .83$ for pure tones, $p < .0001$ for all cases).

We first confirmed the presence of the Bouba–Kiki effect for spoken words (Fig. 6a; mean ± *SD*: 0.12 ± 0.13 for Bouba-like words, 0.58 ± 0.16 for Kiki-like words, $p < .0005$, rank-sum test). Kikiness values in this experiment were highly correlated with those observed in Experiment 2 ($r = .88$, $p < .0005$), thereby reconfirming that Kikiness measured in this manner is a reliable indicator of the strength of the effect.

Next we asked whether pure tones were systematically associated with round and angular shapes as well depending on their frequency. Indeed, the mean Kikiness for the high-frequency was higher compared with low-frequency tones (Fig. 7a; mean ± *SD*: 0.61 ± 0.19 for Bouba-like tones, 0.87 ± 0.11 for Kiki-like tones, $p < .05$). However participants gave higher Kikiness to pure tones compared with the spoken words, suggesting that their responses may not be based on sound frequency alone.

Finally, to understand the audio features that determine the Kikiness, we plotted the Kikiness of each audio stimulus (spoken words and tones) against the mean frequency. This revealed a significant correlation for spoken words ($r = .79$, $p < .0001$; Fig. 7b) and tones ($r = .88$, $p < .0001$; Fig. 7b).

Conclusions

We conclude that the Bouba–Kiki effect is present even for tones with varying frequency: high-frequency tones are associated with spiky shapes and low-frequency tones are associated with rounded shapes.

General discussion

Here, we investigated whether the classic Bouba–Kiki effect is specific to pronounceable words or reflects a more general audio-visual association. To this end, we measured the Bouba–Kiki effect on three types of unpronounceable sounds: reversed spoken words, real object sounds (Experiment 2) and pure tones (Experiment 5). Strikingly, participants systematically associated high-frequency sounds to angular shapes and low-frequency sounds to rounded shapes, regardless of whether they were pronounceable. Pronounceability of a word did not predict the Bouba–Kiki effect, since reversed words showed an equally strong Bouba–Kiki effect despite being less pronounceable (Experiment 3). Finally, mouth shape extracted from video also did not systematically predict the Bouba–Kiki effect (Experiment 4). Sound properties (mean frequency) predicted the Bouba–Kiki effect much better than these speech-related properties. Taken together, these findings show that the Bouba–Kiki effect reflects a general audio-visual association between object shape and sound, rather than any specific association between word names and mouth or speech properties. Below, we review our findings in context of the existing literature.

Our main finding, that the Bouba–Kiki effect is present even for unpronounceable sounds, strongly challenges the mouth-shape hypothesis (Ramachandran & Hubbard, 2001; Sapir, 1929). According to the mouth-shape hypothesis, the mouth makes an angular shape while uttering Kiki-like sounds and a round shape while uttering Bouba-like sounds and this

sensory-motor association gives rise to the Bouba–Kiki effect. If this were true, sounds that cannot be pronounced should have no systematic association with visual shape. Contrary to this prediction, we have found systematic Bouba–Kiki effects for three types of unpronounceable sounds. First, we show that when spoken Bouba- and Kiki-like words are reversed such that they are much less pronounceable without altering their sound frequency content, they are still associated systematically with rounded and angular shapes (Fig. 3). The reversed words are a critical control since they are less pronounceable yet elicited equally strong and systematic associations with shapes (Fig. 4). Second, we show that even real object sounds (Experiment 2) and pure tones (Experiment 5), which have no relation to pronounceability, are also associated systematically with rounded and angular shapes depending on whether they are low or high frequency. Finally, we show that mouth shape, at least as measured using aspect ratio of the mouth, does not systematically predict the Bouba–Kiki effect. Across experiments, the best predictor of the Bouba–Kiki effect was simply the mean frequency of the sound. These findings are inconsistent with the mouth shape hypothesis.

However our findings could be reconciled with the mouth-shape hypothesis by positing that the effect is learned through mouth-shape and sound associations, but is generalized to other sounds, or that the effect when observed for unpronounceable sounds is driven by other mechanisms. Further, it is also possible that articulatory properties other than the mouth aspect ratio (such as tongue movements and oral cavity shape) could explain the Bouba–Kiki effect. We consider these possibilities unlikely, particularly since they cannot be easily disproved.

Instead, we propose that our results support a simpler explanation for the Bouba–Kiki effect—namely, a more general association between sound frequency and shape, which we term the generic shape hypothesis. Participant responses were highly correlated with mean frequency of sounds regardless of their pronounceability, consistent with this possibility. However pure tones, despite having the same mean frequency as spoken words, were more frequently associated with angular shapes (Experiment 5), suggesting that other sound properties may have played a role. We propose that these other properties could be spectral properties (timbre) rather than speech or articulatory properties. These possibilities could potentially be distinguished through cross-language comparisons in which sound features and articulatory features can be decoupled to a degree (Shang & Styles, 2017; Styles & Gawne, 2017; Wong et al., 2022).

Our findings are also consistent with the faster response observed in speeded classification of round/angular shapes when they are paired with low/high frequency tones respectively (Marks, 1987). They are also consistent with the findings that sonorant consonants and voiced bilabial back vowels are associated with round shapes, whereas voiceless stop consonants and unrounded front vowels are associated with angular shapes (D'Onofrio, 2014; Fort et al., 2015; Knoeferle et al., 2017; Westbury et al., 2018). However, the findings of previous studies were also consistent with the possibility that complex speech properties in spoken words could be driving the Bouba–Kiki effect. Our results refute this by showing that the effect is equally strong for reversed words (which contain the same sound frequencies but less pronounceable), and by showing that the effect is predicted by a simple sound property (mean frequency).

Exactly how such an association between sound frequency and shape is learned is unclear. It is certainly plausible that objects with sharp features produce high-frequency sounds, but their material properties also strongly modulate sound frequency (e.g., wooden objects produce low frequency sounds, whereas metallic objects produce higher frequency sounds). We therefore speculate that only certain kinds of audio-visual experience with objects might be required to learn the Bouba–Kiki effect. This is consistent with the recent observation that visual spatial frequencies and auditory temporal frequencies are associated, perhaps through the common modality of touch (Guzman-Martinez et al., 2012). This is also in agreement with previous studies that show that visual experience during development enhances the Bouba–Kiki effect (Fryer et al., 2014; Hamilton-Fletcher et al., 2018).

Alternatively, as alluded earlier, the effect could be initially learned through mouth shape and sound associations but generalized to other sounds later. We consider this unlikely since mouth shape measurements did not predict the Bouba–Kiki effect, and because the effect has been observed even in early childhood (Asano et al., 2015; Imai et al., 2015; Maurer et al., 2006; Ozturk et al., 2013). There could be other explanations for the Bouba–Kiki effect such as shared neural properties (such as larger neural responses for angular shapes and high frequency sounds) or species-general associations between sound frequency and object category (such as low-frequency sounds and threat responses in animals; Sidhu & Pexman, 2018). Our findings do not distinguish between these accounts. Doing so would require characterizing the underlying neural representations or testing these phenomena across species.

One potential limitation of our study is that participants could have known the purpose of this study and might have responded accordingly. We consider this unlikely because, only a relatively small fraction of participants (4% to 21%) indicated after each experiment that they had heard of the Bouba–Kiki effect and our results were unaffected upon excluding them. Such limitations can be overcome using speeded classification tasks where a faster response is observed in classifying a visual stimulus in the presence of an irrelevant auditory stimulus (Marks, 1987; Shang & Styles, 2016; Spence, 2011; L. Walker et al., 2012). We speculate that similar findings would be observed in such implicit association paradigms.

References

Albertazzi, L., Canal, L., & Micciolo, R. (2015). Cross-modal associations between materic painting and classical Spanish music. *Frontiers in Psychology, 6*, 424. https://doi.org/10.3389/fpsyg.2015.00424

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in pre-verbal infants. *Cortex, 63*, 196–205. https://doi.org/10.1016/j.cortex.2014.08.025

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57*(8), 1151–1162. https://doi.org/10.3758/BF03208371

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology, 87*(2), 241–247. https://doi.org/10.1037/h0030524

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition, 126*(2), 165–172. https://doi.org/10.1016/j.cognition.2012.09.007

Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When "Bouba" equals "Kiki": Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports, 6*, 26681. https://doi.org/10.1038/srep26681

Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2019). I know that "Kiki" is angular: The metacognition underlying sound-shape correspondences. *Psychonomic Bulletin & Review, 26*(1), 261–268. https://doi.org/10.3758/s13423-018-1516-8

Chen, Y.-C., Huang, P.-C., & Spence, C. (2021). Global shape perception contributes to crossmodal correspondences. *Journal of Experimental Psychology: Human Perception and Performance, 47*(3), 357–371. https://doi.org/10.1037/xhp0000811

Cowles, J. T. (1935). An experimental study of the pairing of certain auditory and visual stimuli. *Journal of Experimental Psychology, 18*(4), 461–469. https://doi.org/10.1037/h0062202

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the Bouba–Kiki paradigm. *Language and Speech, 57*(3), 367–393. https://doi.org/10.1177/0023830913507694

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology (London, England : 1953), 52*(3), 259–268. https://doi.org/10.1111/j.2044-8295.1961.tb00788.x

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. https://doi.org/10.1016/j.tics.2015.07.013

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 6.1–12. https://doi.org/10.1167/10.1.6

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are more important than vowels in the Bouba–Kiki effect. *Language and Speech, 58*(Pt 2), 247–266. https://doi.org/10.1177/0023830914534951

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic-auditory associations in the "Bouba–Kiki" effect. *Cognition, 132*(2), 164–173. https://doi.org/10.1016/j.cognition.2014.03.015

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191–1203. https://doi.org/10.3758/bf03193720

Gold, R., & Segal, O. (2017). The Bouba–Kiki effect and its relation to the Autism Quotient (AQ) in autistic adolescents. *Research in Developmental Disabilities, 71*, 11–17. https://doi.org/10.1016/j.ridd.2017.09.017

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology, 22*(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition, 175*, 114–121. https://doi.org/10.1016/j.cognition.2018.02.023

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology, 109*(2), 219–238. https://doi.org/10.2307/1423274

Hung, S. M., Styles, S. J., & Hsieh, P. J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science, 28*(3), 263–275. https://doi.org/10.1177/0956797616677313

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLOS ONE, 10*(2), e0116494. https://doi.org/10.1371/journal.pone.0116494

Karthikeyan, S., Rammairone, B., & Ramachandra, V. (2016). The Bouba–Kiki phenomenon tested via schematic drawings of facial expressions: Further validation of the internal simulation hypothesis. *I-Perception, 7*(1), 2041669516631877. https://doi.org/10.1177/2041669516631877

Kim, S.-H. (2020). Bouba and Kiki inside objects: Sound-shape correspondence for objects with a hole. *Cognition, 195*, 104132. https://doi.org/10.1016/j.cognition.2019.104132

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), 5562. https://doi.org/10.1038/s41598-017-05965-y

Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung, 31*(1), 18–30. https://doi.org/10.1007/BF00422382

Ković, V., Sučević, J., & Styles, S. J. (2017). To call a cloud "cirrus": Sound symbolism in names for categories or items. *PeerJ, 2017*(6), 1–18. https://doi.org/10.7717/peerj.3466

Liew, K., Lindborg, P. M., Rodrigues, R., & Styles, S. J. (2018). Cross-modal perception of noise-in-music: Audiences generate spiky shapes in response to auditory roughness in a novel electroacoustic concert setting. *Frontiers in Psychology, 9*, 1–12. https://doi.org/10.3389/fpsyg.2018.00178

Lim, J. D. F., & Styles, S. J. (2016). Super-normal integration of sound and vision in performance. *Array. The Journal of the ICMA, 2016*(August 2015), 45–49. https://doi.org/10.25370/array.v20152523

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology, 6*, 1246. https://doi.org/10.3389/fpsyg.2015.01246

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences of the United States of America, 108*(51), 20661–20665. https://doi.org/10.1073/pnas.1112605108

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 384–394. https://doi.org/10.1037//0096-1523.13.3.384

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9*(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

O'Boyle, M. W., & Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research, 9*(6), 535–544. https://doi.org/10.1007/BF01068115

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience, 3*(3/4), 348–355. https://doi.org/10.1080/17470910701563681

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The Takete—Maluma phenomenon in autism spectrum disorders. *Perception, 42*(2), 233–241. https://doi.org/10.1068/p7357

Owens, A., Isola, P., McDermott, J., Torralba, A., Adelson, E. H., & Freeman, W. T. (2016). Visually Indicated Sounds. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016-Decem*, 2405–2413. https://doi.org/10.1109/CVPR.2016.264

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology, 114*(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*(3/4), 319–333. https://doi.org/10.1007/s00221-012-3140-6

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

Rogers, S. K., & Ross, A. S. (1975). A cross-cultural test of the Maluma–Takete phenomenon. *Perception, 4*(1), 105–106. https://doi.org/10.1068/p040105

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*(3), 225–239. https://doi.org/10.1037/h0070931

Shang, N., & Styles, S. J. (2016). An Implicit Association Test on audiovisual crossmodal correspondences. *Array. The Journal of the ICMA, 2016*(August 2015), 50–51. https://doi.org/10.25370/array.v20152524

Shang, N., & Styles, S. J. (2017). Is a High Tone Pointy? Speakers of Different Languages Match Mandarin Chinese Tones to Visual Shapes Differently. *Frontiers in Psychology, 8*, 1–13. https://doi.org/10.3389/fpsyg.2017.02139

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review, 25*(5), 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Sievers, B., Lee, C., Haslett, W., & Wheatley, T. (2019). A multi-sensory code for emotional arousal. *Proceedings. Biological Sciences, 286*(1906), 20190513. https://doi.org/10.1098/rspb.2019.0513

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Styles, S. J., & Gawne, L. (2017). When does Maluma/Takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *I-Perception, 8*(4), 204166951772480. https://doi.org/10.1177/2041669517724807

Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape-sound-meaning congruence in words, and shape-sound congruence in pseudowords. *Brain and Language, 145–146*, 11–22. https://doi.org/10.1016/j.bandl.2015.03.011

Turoman, N., & Styles, S. J. (2017). Glyph guessing for 'oo' and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society Open Science, 4*(9), 170882. https://doi.org/10.1098/rsos.170882

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41*(10), 1186–1192. https://doi.org/10.1068/p7149

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language, 99*, 122–150. https://doi.org/10.1016/j.jml.2017.09.006

Wong, L. S., Kwon, J., Zheng, Z., Styles, S. J., Sakamoto, M., & Kitada, R. (2022). Japanese sound-symbolic words for representing the hardness of an object are judged similarly by Japanese and English speakers. *Frontiers in Psychology, 13*, 830306-830306. https://doi.org/10.3389/fpsyg.2022.830306