



HAL
open science

The cerebral bases of the bouba-kiki effect

Nathan Peiffer-Smadja, Laurent Cohen

► **To cite this version:**

Nathan Peiffer-Smadja, Laurent Cohen. The cerebral bases of the bouba-kiki effect. *NeuroImage*, 2019, 186, pp.679-689. 10.1016/j.neuroimage.2018.11.033 . hal-02003299

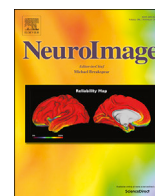
HAL Id: hal-02003299

<https://hal.sorbonne-universite.fr/hal-02003299v1>

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



The cerebral bases of the bouba-kiki effect

Nathan Peiffer-Smadja^a, Laurent Cohen^{a,b,*}

^a *Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, F-75013, Paris, France*

^b *Département de Neurologie 1, Hôpital de la Pitié Salpêtrière, AP-HP, F-75013, Paris, France*



ARTICLE INFO

Keywords:

Crossmodal
Speech perception
Shape
Language
fMRI

ABSTRACT

The crossmodal correspondence between some speech sounds and some geometrical shapes, known as the bouba-kiki (BK) effect, constitutes a remarkable exception to the general arbitrariness of the links between word meaning and word sounds. We have analyzed the association of shapes and sounds in order to determine whether it occurs at a perceptual or at a decisional level, and whether it takes place in sensory cortices or in supramodal regions. First, using an Implicit Association Test (IAT), we have shown that the BK effect may occur without participants making any explicit decision relative to sound-shape associations. Second, looking for the brain correlates of implicit BK matching, we have found that intermodal matching influences activations in both auditory and visual sensory cortices. Moreover, we found stronger prefrontal activation to mismatching than to matching stimuli, presumably reflecting a modulation of executive processes by crossmodal correspondence. Thus, through its roots in the physiology of object categorization and crossmodal matching, the BK effect provides a unique insight into some non-linguistic components of word formation.

1. Introduction

In 1929, Wolfgang Köhler showed participants two novel shapes, one roundish and one spiky, and proposed the two pseudowords “baluma” (changed into “maluma” in later studies) and “takete” (Köhler, 1929, 1947). Participants were simply asked to decide which pseudoword would more naturally match which shape. Köhler (1947) observed that “most people answered without any hesitation”, choosing “b/maluma” for the round shape and “takete” for the spiky shape. This simple experiment brought to light a strong and consensual link (Chen et al., 2018) between meaningless speech sounds and geometrical shapes, an effect often referred to as the bouba-kiki effect (BK; Ramachandran and Hubbard, 2001).

The BK effect is one among a multitude of crossmodal correspondences (for a review, see Spence, 2011), including links between different modalities or conceptual fields, such as for instance “down” being associated with darkness, small numbers, low-pitch sounds, sadness, as opposed to “up” being associated with light, large numbers, high-pitch sounds, joy, etc. Strictly speaking, the term crossmodal correspondence should apply only to compatibility effects between attributes of a stimulus which are perceived through different sensory modalities (including the BK effect) (see Spence, 2011 for terminological clarifications). However, it has occasionally been extended to more

conceptual-sensory associations, e.g. between facial emotional expressions and colors (Palmer et al., 2013). It has been argued that full-blown synaesthesia may represent extreme forms of such common crossmodal correspondences (Martino and Marks, 2001). Indeed, it has been shown that synaesthetes show stronger BK pseudoword-shape correspondences (but not more basic correspondences) than non-synaesthetes (Lacey et al., 2016). However the links between synaesthesia and mere correspondences remain a controversial issue (Deroy and Spence, 2013).

Among crossmodal correspondences, the core specificity of the BK effect is that it involves speech sounds. It may therefore be relevant to the link between word sound and word meaning, an issue which has been scrutinized throughout the history of science and philosophy, from Plato's Cratylus and medieval scholiasts to Saussure's work and contemporary cognitive neuroscience. Basically, this link is considered arbitrary, with however some qualifications. Thus onomatopoeia features acoustic similarity between speech patterns and the objects referred to (e.g. bees are buzzing) (Taitz et al., 2018). At a more abstract level, plurality may be expressed by means of word duplication (e.g. teman-teman referring to the plural “some friends” in Indonesian). Such systematic links may explain why the meaning of some unknown foreign words can be guessed better than chance by naïve monolingual participants (Reville et al., 2014). The BK effect is a further type of

* Corresponding author. Institut du Cerveau et de la Moelle épinière 47 Boulevard Hôpital, 75013 Paris, France.

E-mail address: laurentcohen2@gmail.com (L. Cohen).

<https://doi.org/10.1016/j.neuroimage.2018.11.033>

Received 19 August 2018; Received in revised form 18 November 2018; Accepted 21 November 2018

Available online 29 November 2018

1053-8119/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

deviation from the arbitrariness of sound-meaning relationships, based upon crossmodal correspondences.

Previous research has clarified some features accounting for the BK effect. Thus, consonants have a greater influence than vowels (Fort et al., 2015); vowel backness, consonant voicing, and consonant place of articulation all elicit additive effects (D'Onofrio, 2014). Knoeferle et al. (2017) assessed the role of vowel formants by asking participants to evaluate the subjective visual “size” and “shape” of simple syllables. They showed that size judgments were predicted by the first and second formants, and by vowel duration, while shape judgments were predicted by the second and third formants. Moreover, although the BK effect prevails in cultures with no written language (Bremner et al., 2013), and among pre-reader children (Maurer et al., 2006), it shows subtle modulations according to cultural factors. Thus, the curvature of letter shapes modulates the BK effect (Cuskey et al., 2017). Disentangling spatial frequency, spatial amplitude, and spikiness in the design of visual shapes, Chen et al. (2016) showed that while both groups evinced a strong BK effect, North Americans were more sensitive to amplitude, but less sensitive to spikiness, when compared with Taiwanese participants.

Most accounts of the BK effect relate it to statistical correlations between matching sounds and shapes. Such correlations may occur in the external world. Thus, the crossmodal correspondence between high- vs low-pitch sounds and small vs large objects may be attributed to the fact that smaller objects tend to resonate at higher frequencies than larger objects. Similarly, harder objects also tend to resonate at higher frequencies (kiki) and break into sharper pieces (spiky) than softer objects, which resonate at lower frequencies (bouba) and assume rounder shapes (Parise and Spence, 2012). Sound-shape correlations have also been claimed to occur within the speech processing system. Thus Ramachandran and Hubbard (2001) suggested that the sharp changes in the visual direction of lines in spiky shapes mimic the sharp phonemic inflections of the sound kiki, but also the sharp inflections of the tongue on the palate.

Whatever its auditory-visual and motor-visual origins, the perceived correspondence between round/spiky shapes and bouba/kiki sounds reflects greater crossmodal integration of auditory and visual features for the categorization of multimodal objects (Bizley et al., 2016). Once framed in the context of multisensory integration, trying to bring to light the brain mechanisms of the BK effect raises two related issues.

First, does the integration of shapes and sounds occur at a perceptual/automatic or at a decisional/controlled level (for a critical discussion, see Spence and Deroy, 2013)? Evans and Treisman (2010) studied the correspondence between auditory pitch and visual elevation. They showed that matching stimuli yielded better performance in unimodal judgement tasks, irrespective of whether the task did or did not involve the dimensions of pitch and elevation, suggesting that effects reflected automatic perceptual binding rather than decision-related factors. Still, the conclusion that task set is irrelevant to the integration of shapes and sounds cannot be easily generalized. Thus, Chiou and Rich (2012) confirmed that auditory cues with a matching pitch facilitated decision on the elevation of visual targets. However, this effect did not interact with the delay between target and cue. Moreover, it could be reversed when participants expected targets to appear at a location that was opposed to the spontaneous mapping of pitch and elevation. However, the reversed congruency effect was sustained across a longer range of delays than the default congruency effect. This pattern suggests that the correspondence was influenced by relatively late and controlled decisional processes.

Second, in the BK effect, does the integration take place in sensory cortices or in supramodal regions? In animals and humans, multisensory interactions have been found at essentially all levels of the central nervous system (Alais et al., 2010). For instance, Sadaghiani et al. (2009) biased visual motion perception using 3 types of auditory cues with an increasing cultural weight, involving actual movement in space, metaphorical up/down pitch movement, or verbal left/right labels. They showed that these 3 types of cues had a decreasing influence in the

audiovisual motion area MT, but an increasing influence in the right IPS, a higher-level convergence region (see also Bien et al., 2012).

Research on multisensory integration in speech perception mostly studied the integration of lip reading cues with auditory speech signals (for a review, see Kilian-Hütten et al., 2017). Evidence of such integration has been found in higher-order supramodal areas, including the posterior superior temporal sulcus (pSTS; see e.g. Calvert, 2001) and the prefrontal cortex (Nath and Beauchamp, 2012; Ojanen et al., 2005). It has also been found in early auditory areas generally considered unimodal (Ghazanfar and Schroeder, 2006; van Wassenhove et al., 2005).

Once framed in this context, we predict that the BK effect should have both a perceptual (and possibly more automatic) component, whereby sensory processing in the auditory or visual cortices would be facilitated by crossmodal matching, and a decisional component whereby prefrontal areas would be sensitive to crossmodal matching in a task-dependent manner. In the present study of the BK effect, we address the two questions of automaticity and brain localization. We first validate the basic BK phenomenon with an explicit task, allowing us to identify the best experimental parameters for the following experiments. We then study whether the BK effect still prevails with an implicit task derived from the Implicit Association Test (IAT) paradigm (Greenwald et al., 1998), as first assessed by Parise and Spence (2012) and Lacey et al. (2016), a task in which participants do not respond based on their introspection, and are not informed that the topic of the study is crossmodal matching. This also allows us to extract an individual index of sensitivity to the implicit BK effect. These two behavioral experiments set the stage for an fMRI study in which we identify the cerebral correlates of the implicit BK effect, both in sensory and in supramodal cortical regions.

2. Behavioral study of explicit sound-shape associations

The aims of this experiment were (a) to confirm the basic BK phenomenon, i.e. that some pseudowords are more tightly associated to particular shapes and sizes, (b) to study the respective role of vowels and consonants in such linkage, and (c) to select optimal stimuli for the following experiments.

2.1. Methods

2.1.1. Material

We designed a set of 4 round and 4 spiky stimuli. For each stimulus, a large and a small version were derived, resulting in four types of stimuli (large/small size x round/spiky shape type) (Fig. 1a). Large stimuli were fitted in a rectangle of $17.2^\circ \times 11.5^\circ$, and small stimuli in a rectangle of $10.3^\circ \times 6.9^\circ$. Stimuli were light gray on a black background.

We created 24 pseudowords with a CVCV structure, varying along two dimensions: vowel type (front unrounded [i] and [e] vs. back rounded [u] and [o]), and consonant type (unvoiced stops [p], [k] and [t] vs. voiced continuants [l], [m] and [j]), yielding four types of pseudowords. The set of stimuli was created by selecting among all possible combinations, many of which were real words, 6 pseudowords for each of the 4 conditions of vowel type x consonant type (e.g. “keti”, “toku”, “lije”, “mujo”).

2.1.2. Procedure

During each trial, participants listened to a pseudoword over headphones, immediately followed by two visual stimuli shown side by side. These differed in shape type, or in size, or both. Participants had to choose which visual stimulus was matching the pseudoword best, and answer by pressing a button with their right or left hand.

Each of the 24 pseudo-words was associated with the six possible combinations of shape type and size (large round and large spiky, large round and small spiky, large round and small round, large spiky and small round, small spiky and small round), yielding a total of 144 trials per participant. For each trial, visual stimuli were picked randomly out of the appropriate set. Participants were

Explicit sound-shape associations

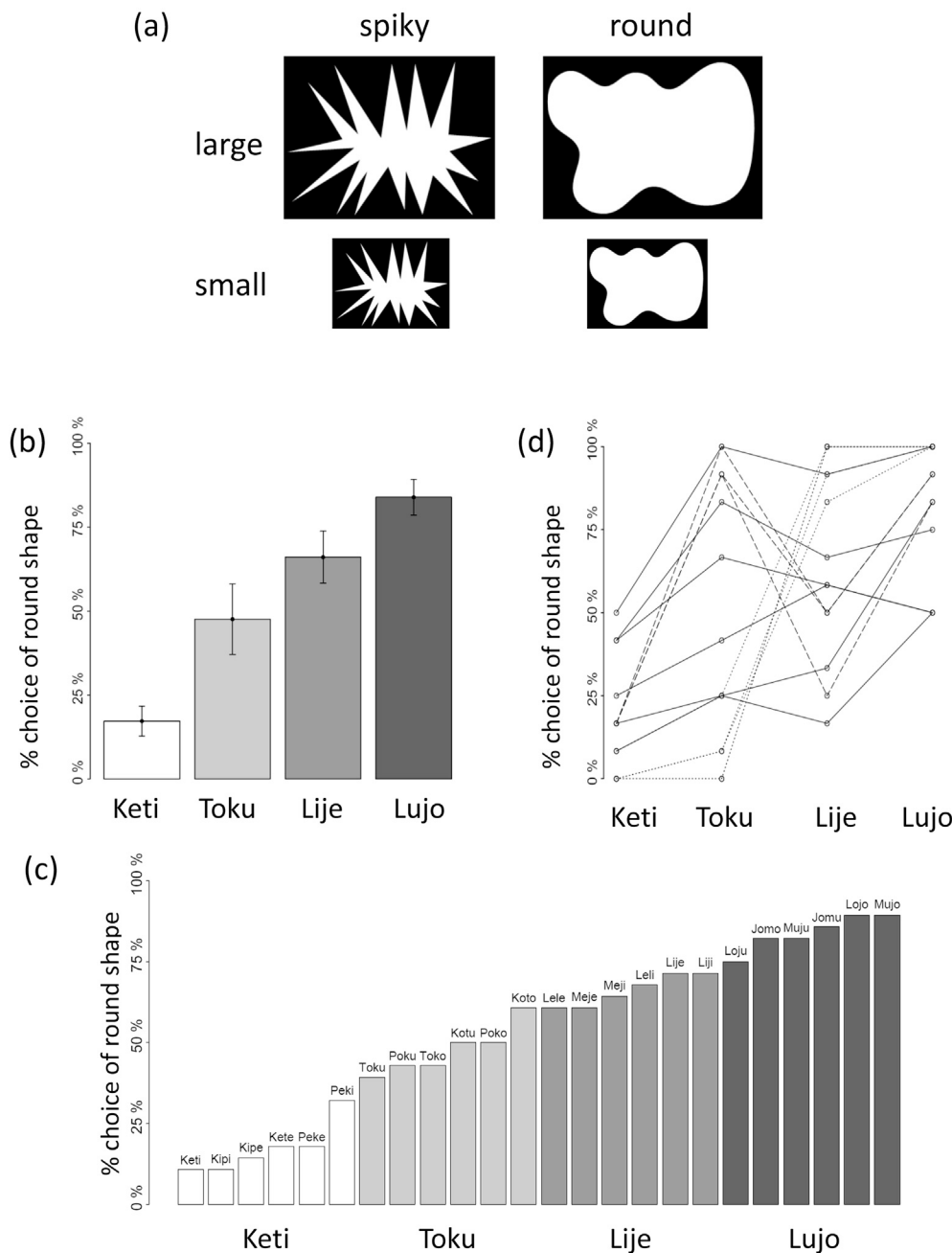


Fig. 1. Results of the experiment on explicit sound-shape associations. (a) Examples of visual stimuli, crossing shape and size. Similar shapes were used in the following experiments. Panels (b)–(d) display the average percent choice of the round over the spiky shape to match to 4 proposed types of pseudowords. Ketu: stop unvoiced consonants with front unrounded vowels; Toku: stop unvoiced consonants with back rounded vowels; Lije: continuant voiced consonants with front unrounded vowels; Lujo: continuant voiced consonants with back rounded vowels. (b) Round shapes were more often associated to continuant voiced than unvoiced stop consonants, and to back rounded than front unrounded vowels. Error bars represent ± 1 SEM. (c) This pattern of cross-modal association was highly consistent across all stimuli. Results are displayed for all pseudowords, ordered by increasing rate of choice of round shapes. (d) Each line represents one subject. The relative impact of vowels and consonants on crossmodal associations differed across participants, ranging from a predominant effect of vowels (dashed lines) to a predominant effect of consonants (dotted lines), with about one half of participants showing intermediate patterns (solid lines).

explicitly told that there was no determined correct answer, and that they had to follow their intuition, with no time limit.

2.1.3. Participants

Fourteen right-handed native French speakers, 19–49 years old (8 women, mean age: 31 years), with normal hearing and normal or corrected-to-normal vision, took part in this experiment. Handedness was assessed with the Edinburgh Handedness Inventory (mean score: +70, SD: 21). The project was approved by the regional ethical committee and the participants gave their written informed consent.

For each participant and each type of pseudowords, we computed the percentage of choice of a round vs a spiky, and of a large vs a small visual stimulus. These values were entered in two ANOVAs, in order to study

preference for shape type and for size. Each model included two within-participants factors (vowel type and consonant type) and one random factor (participants). In the analysis of shape preference, we only took into account trials wherein the two visual stimuli did not differ in size; in the analysis of size preference, we only took into account trials wherein the two visual stimuli did not differ in shape type.

2.2. Results

2.2.1. Shape preference

Back rounded vowels [u] and [o] were more associated to round shapes, and front unrounded vowels [i] and [e] to spiky shapes (65.8% and 41.7% choice of the round shape, respectively; $F(1,13) = 15.07$,

$p < 0.002$) (Fig. 1b). Voiced continuant consonants [l], [m] and [j] were more associated to round shapes, and unvoiced stop consonants [p], [k] and [t] to spiky shapes (75% and 32.4% choice of the round shape, respectively; $F(1,13) = 15.77$, $p < 0.002$). The effect of consonant type was larger than the effect of vowel type, as participants more often chose a round shape for lije-type pseudowords than for toku-type pseudowords, the two cases in which vowels and consonants pulled responses in divergent directions. There was no interaction of vowel type and consonant type ($F(1,13) = 3.74$, $p > 0.05$).

To check whether these sound-shape associations were consistent across items, or due to a subset of pseudowords or phonemes, we plotted for each of the 24 pseudowords the percent choice of the round shape. As shown in Fig. 1c, pseudowords ranked perfectly according to the types of vowel and consonant, demonstrating the robustness of this effect across items.

In order to qualitatively assess whether sound-shape associations were consistent across participants, we plotted the percentage of shape choice of the 14 participants for the four classes of pseudowords. In Fig. 1d, each line represents one participant's performance. Three participants appeared mostly sensitive to the vowel effect (Fig. 1d, dashed lines), 4 to the consonant effect (Fig. 1d, dotted lines), while the other half of the group had a less contrasted profile (Fig. 1d, solid lines).

2.2.2. Size preference

Back rounded vowels were more associated to large visual stimuli than front unrounded vowels, and conversely (66.1% and 48.2% choice of the large stimulus for back and front vowels, respectively; $F(1,13) = 10.35$, $p < 0.01$). There was no significant effect of consonant type on size preference ($F(1,13) = 0.30$, $p = 0.6$), and no interaction of vowel type and consonant type ($F(1,13) = 0.56$, $p = 0.5$).

2.3. Summary

First, we demonstrated a strong association between shape and sound. Front unrounded vowels and unvoiced stop consonants were associated to spiky shapes, while back rounded vowels and voiced continuant consonants were associated to round shapes, in agreement with previous research (D'Onofrio, 2014; Fort et al., 2015; Köhler, 1947). Second, the relative impact of vowel type and of consonant type differed across participants. Third, the association between size and sound was weaker, and concerned only vowels, as shown by earlier studies (Tarte and Barritt, 1971). We concluded that in subsequent experiments, in order to observe robust effects of association, we should manipulate shape rather than size, and associate rather than separate the effects of vowel and consonant types.

3. Behavioral study of implicit sound-shape associations

Once we had established the experimental parameters most appropriate for eliciting a BK effect, we moved to the core topic of this study, that is to say the exploration of implicit sound-shape association. A set of participants first participated in a behavioral experiment, and then in an fMRI study. The aims of the behavioral experiment were, first, to assess whether participants would show a Bouba-Kiki effect even when no explicit judgement was required on audio-visual matching; and, second, to obtain an individual index of sensitivity to this effect, for use in the subsequent fMRI experiment.

In order to determine whether a BK effect may occur without participants making any explicit decision relative to sound-shape associations, we used a variant of the Implicit Association Test (IAT) (Greenwald et al., 1998). This method has been mostly used in social psychology to observe effects that are not found by explicit questioning, for example to assess racial prejudices which participants would not overtly confess (Greenwald and Banaji, 2017; Phelps et al., 2000; Xu et al., 2014). The IAT provides a measure of the association between two pairs of contrasted concepts (e.g. flowers vs insects, and pleasant vs unpleasant).

Participants have to categorize words into one of the four combinations defined by the two pairs of concepts. The critical trick of the IAT is that responses are faster and more accurate when concepts that are strongly associated share the same behavioral response, e.g. if responses to names of insects and to unpleasant words should be produced using the same hand. In the present experiment, the two pairs of concepts were round vs spiky shapes and bouba vs kiki sounds, which should allow us to assess association between sounds and shapes, without requiring any explicit choice. We predicted that responses would be faster and more accurate in congruent blocks, i.e. whenever kiki sounds and spiky shapes (and bouba sounds and round shapes) should be classified using the same hand.

3.1. Methods

3.1.1. Material

We used two classes of bisyllabic CVCV pseudowords selected from the material of the previous experiment: pseudowords with voiced continuant consonants [l], [m], [j] and back rounded vowels including at least one [o] (eg “moju”, henceforth called “bouba pseudowords”), which were proved to be associated to round shapes, and pseudowords with unvoiced stop consonants [p], [k], [t] and frontal unrounded vowels including at least one [i] (eg “kipe”, henceforth called “kiki pseudowords”), which were associated to spiky shapes. We used a total of 24 pseudowords with front vowels and stop unvoiced consonants, and 24 pseudowords with back vowels and continuant voiced consonants. We also selected 18 spiky shapes and 18 round shapes.

3.1.2. Procedure

For each trial, participants were simultaneously presented with a pseudoword and a shape. For half of the trials the pseudoword and shape were matching, and for the other half they were mismatching. Participants had to perform a double classification task. First, they had to decide if the pseudoword contained the sound “o” or the sound “i”. Then they had to decide if the shape was round or spiky. They responded by pressing a left-hand or a right-hand button according to instructions (see below). The response to pseudowords was prompted by a loudspeaker icon appearing on the screen, and the response to shapes was prompted by the icon of an eye. Stimuli appeared for 600 ms, followed by the loudspeaker icon. As soon as the first answer was given, the loudspeaker icon was replaced by the icon of the eye. A maximum duration of 1500 ms was allowed for each response. As soon as the second answer was given, the icon of the eye was replaced by a central fixation cross, which remained visible for 3 s minus the sum of the two response times, thus yielding a constant SOA of 3600 ms. The experiment was divided into 4 blocks, each comprising 80 trials plus 10 initial training trials. For training trials, participants had unlimited time to respond to both the pseudoword and the shape.

Crucially, instructions changed across blocks. In the two congruent blocks, participants had to answer to round shapes and bouba pseudowords with one hand, and to spiky shapes and kiki pseudowords with the other hand. Conversely, in the two incongruent blocks, participants had to answer to spiky shapes and bouba pseudowords with one hand, and to round shapes and kiki pseudowords with the other hand. In all blocks, matching and mismatching pairs were presented in equal proportion. For half the participants, the order of the sessions was congruent-incongruent-incongruent-congruent (CIIC) and for the other half the order was ICCI. The hand-sound pairing (e.g. left-bouba, right-kiki) remained the same through the whole experiment, and hence the hand-shape pairing was inverted between congruent and incongruent blocks. The hands associated to bouba and kiki pseudowords were counter-balanced across participants.

3.1.3. Participants

Eighteen native right-handed French speakers, 19–35 years old (9 men, mean age: 23 years), with normal hearing and normal or corrected-to-normal vision, participated in the study. They had not participated in

the previous experiment. Handedness was assessed with the Edinburgh Handedness Inventory (mean score: +73, SD: 19). The project was approved by the regional ethical committee and the participants gave their written informed consent.

3.2. Results

Error rates and mean correct RTs (measured from stimulus onset) were computed for each participant and for each condition, and entered in ANOVAs with congruence and matching as within-participant factors and participants as random factor. ANOVAs were performed for responses to pseudowords and responses to shapes (Fig. 2).

Error rates were 6% and 5%, and mean RTs were 532 ms and 417 ms, for responses to pseudowords and shapes, respectively. There was a main effect of congruence, as congruent blocks yielded lower error rates and faster responses than incongruent blocks, for both pseudowords (mean error rate: 3.0% vs 8.2%; mean RT: 472 ms vs 607 ms) and shapes (mean error rate: 2.9% vs 7.0%; mean RT: 377 ms vs 466 ms). The effect was

significant in all 4 ANOVAs ($p < 0.001$ for RTs, and $p < 0.02$ for errors). The effect of congruence on error rates was confirmed using the non-parametric Wilcoxon signed rank test, for responses both to pseudowords ($p < 0.001$) and to shapes ($p = 0.01$). For the effect of congruence, we computed Cohen's d-score, a standardized measure of effect size, finding values of 1.15 and 0.68 for RTs and error rates, respectively, roughly corresponding to "large" effect sizes (Sawilowsky, 2009). There was also a significant interaction of congruence x matching (in all 4 ANOVAs: $p < 0.02$ for both RTs and errors). This interaction simply reflects an effect of same/different response hands: matching trials in congruent blocks and mismatching trials in incongruent blocks were easier because the same hand was used to classify both the pseudoword and the shape.

Finally, we computed for each participant the difference in error rate for incongruent minus congruent blocks, and used this individual index of sensitivity to implicit sound-shape association as a regressor in subsequent analyses of fMRI data.

Implicit sound-shape associations

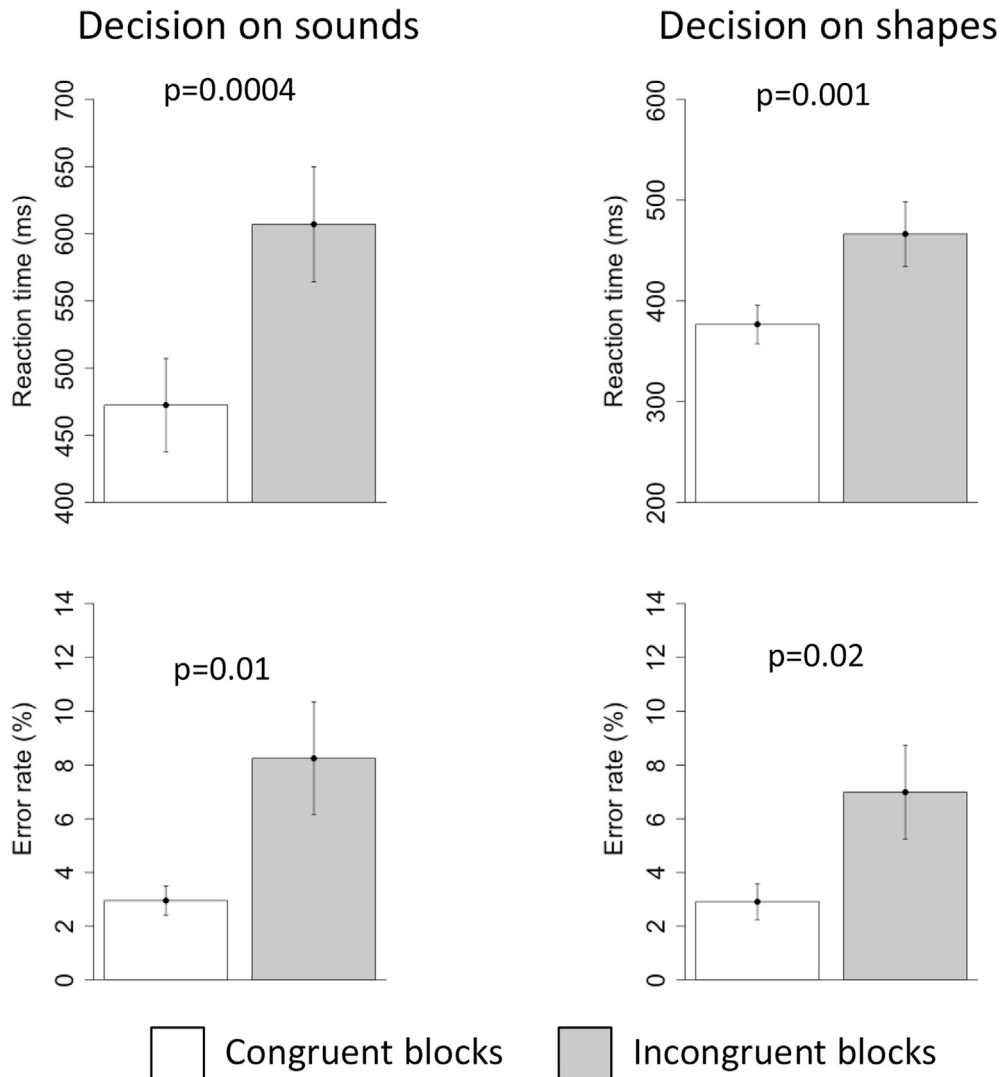


Fig. 2. Results of the experiment on implicit sound-shape associations. In an Implicit Association Test (IAT), responses were faster and more accurate in blocks where responses to shapes and sounds were congruent than in blocks where they were incongruent. Error bars represent ± 1 SEM.

3.3. Summary

As predicted, responses were faster and more accurate in congruent than in incongruent blocks, demonstrating that the bouba-kiki sound-shape association had an impact on behavior even when it was irrelevant to the task and not even mentioned to participants as a parameter of interest.

4. fMRI study of implicit sound-shape associations

4.1. Methods

4.1.1. Material

We used the same auditory and visual stimuli as in the previous experiment, except that we also included pseudowords with two [u] or with two [e], which were discarded from the behavioral experiment due to task design.

4.1.2. Procedure

Pseudowords and shapes were used to create 4 basic types of experimental blocks: auditory blocks consisting of pseudowords, visual blocks consisting of shapes, bimodal matching blocks consisting of matching pseudoword-shape pairs, and bimodal mismatching blocks consisting of mismatching pseudoword-shape pairs. Each basic type of block existed in two variants, yielding a total of 8 types of blocks: auditory blocks comprised either bouba or kiki pseudo words, visual blocks either round or spiky shapes, matching blocks either bouba-round or kiki-spiky pairs, and mismatching blocks either bouba-spiky or kiki-round pairs. Ten blocks of each type were shown, plus 20 resting blocks of the same duration as activation blocks. Within each block, stimuli were selected randomly from the appropriate category of items. Blocks were mixed in a pseudo-random order, different in all participants. Each block included 10 trials, consisting of a central fixation point (100 ms), followed by the stimulus (600 ms, i.e. the duration of the longest pseudoword). Shapes were shown in light gray on a black background, simultaneously with the onset of pseudowords on bimodal trials.

Participants were simply asked to pay attention to both visual and auditory stimuli, and to detect occasional targets. Targets consisted either of a cross shown at fixation point, or of a short “beep” sound. Targets randomly replaced one out of 20 stimuli. In visual blocks, targets were crosses, in auditory blocks targets were “beeps”, and in bimodal blocks either crosses or “beeps”. Participants had to respond to targets by pushing a button with their right hand, as quickly and as accurately as possible.

4.1.3. Acquisition parameters

We used a 3-T MRI (Siemens Trio TIM) with a 12-channel head coil, and a gradient-echo planar imaging sequence sensitive to brain oxygen-level dependent (BOLD) contrast (40 contiguous axial slices, acquired using ascending interleaved sequence, 3 mm thickness; TR = 2400 ms; flip angle = 90°, TE = 30 ms, in-plane resolution = 3 × 3 mm, matrix = 64 × 64). For each acquisition, the first 4 vol were discarded to reach equilibrium. T1-weighted images were also acquired for anatomical localization. We acquired a total of 301 functional volumes.

4.1.4. Statistical analysis

Individual data processing, performed with SPM8 software, included corrections for EPI distortion, slice acquisition time, and motion; normalization to the MNI anatomical template; Gaussian smoothing (5 mm FWHM); and fitting with a linear combination of functions derived by convolving the time series of events with the standard hemodynamic response function implemented in the SPM8 software (a combination of 2 gamma functions, with a rise peaking around 6 s followed by a longer undershoot), without including in the model the temporal derivatives of these functions. There was thus a total of 10 regressors (8 types of trials plus 2 types of targets). Individual contrast images were computed for

each stimulus type minus baseline, then smoothed (5 mm FWHM), and eventually entered in an ANOVA for random effect group analysis. We used a voxelwise threshold of $p < 0.001$ for effects of modality, type of sound and type of shape, and or of $p < 0.01$ for smaller effects involving intermodal matching, always with a threshold for cluster extent of $q_{FDR\ corr} < 0.05$. Unless stated otherwise, results were corrected for multiple comparisons across the whole brain volume. Whenever results were corrected within a region of interest (ROI), the ROI was defined using orthogonal contrasts, in order to avoid “double dipping” or statistical circularity (Friston et al., 2006; Kriegeskorte et al., 2009; Poldrack et al., 2008). In order to find voxels whose activation was correlated across participants with the behavioral score, we also entered individual images of contrasts of interest in linear regressions with behavioral scores.

4.1.5. Participants

The 18 participants were the same as in the previous behavioral experiment.

4.2. Behavioral results

Participants had to detect occasional visual or auditory targets. The mean detection rate was 89% and 87%, respectively. The overall detection rate ranged from 71% to 100% across participants, except for one participant who scored below 50% and was excluded from subsequent analyses. We also excluded one participant who did not hear auditory stimuli due to technical malfunction.

4.3. Imaging results

4.3.1. Effects of sensory modality

Relative to rest, auditory stimuli activated the superior temporal gyrus and sulcus bilaterally (left: MNI -60 -28 6, $Z > 8$; right: MNI 66 -20 2, $Z > 8$), and superior precentral cortex bilaterally (left: MNI -48 4 50, $Z = 6.04$; right: MNI 56 0 44, $Z = 5.67$) (Fig. 3a). Visual stimuli activated the occipital cortex (left: MNI -10 -102 6, $Z > 8$; right: 28 -92 16, $Z > 8$), the ventral temporal cortex posterior to about $y = -35$, and the intraparietal cortex bilaterally (left: MNI -28 -56 58, $Z = 4.59$; right: MNI 30 -54 52, $Z = 4.71$). Bimodal stimuli activated the union of the two unimodal sets of regions.

4.3.2. Effects of shape and sound

We pooled unimodal and bimodal trials and compared spiky vs round shapes and bouba vs kiki sounds, correcting the statistics within the volumes activated by visual and auditory stimuli minus rest, respectively.

The contrast of round minus spiky shapes activated the anterior calcarine cortex (left: MNI -10 -82 8, $Z = 5.61$; right: MNI: 6 -74 14, $Z = 6.46$) (Fig. 3b). Conversely, spiky minus round shapes activated the bilateral posterior occipital lobes, including the posterior calcarine cortex (left: MNI -12 -94 -8, $Z > 8$; right: MNI 14 -98 -6, $Z > 8$). These differences resulted from the more peripheral and more central average stimulation for round and spiky shapes, respectively.

The contrast of bouba minus kiki pseudowords activated a region in the left superior temporal gyrus overlapping with Heschl's gyrus as defined by the AAL atlas (Tzourio-Mazoyer et al., 2002) (MNI -50 -10 4, $Z = 5.99$) (Fig. 3c). There was a symmetrical right-hemispheric activation somewhat below the threshold for cluster extent (MNI 52 -8 2, $Z = 4.58$, 137 voxels). The opposite contrast (kiki minus bouba pseudowords) did not yield significant activations.

4.3.3. Effects of intermodal matching

The contrast of mismatching minus matching bimodal stimuli showed bilateral dorsolateral prefrontal activation (left: MNI -42 44 22, $Z = 3.92$; right: MNI 34 56 28, $Z = 3.50$) in the inferior, middle and superior frontal gyri (Fig. 4). The opposite contrast (matching minus mismatching stimuli) did not yield significant differences.

We then explored whether in sensory regions sensitive to the

Effects of modality, shape, and sound

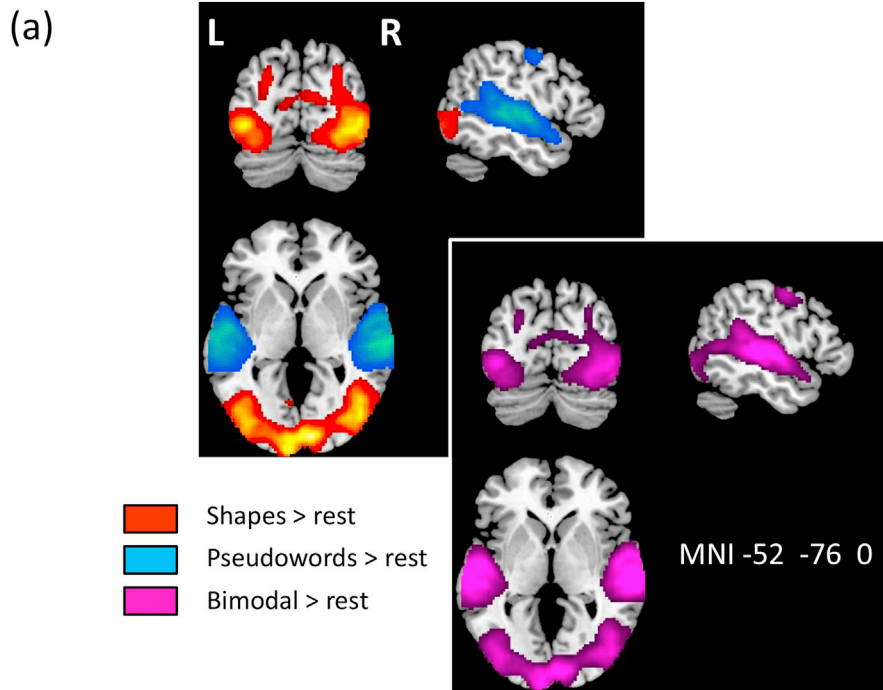
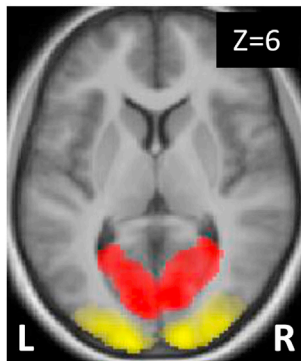
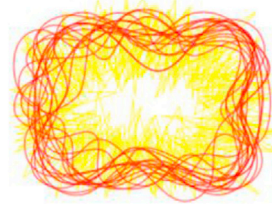


Fig. 3. (a) Brain activations by visual (red), auditory (blue), and bimodal (purple) stimuli relative to rest. (b) Activations of the occipital cortex to round more than to spiky shapes (red), and the opposite contrast (yellow), reflecting differences in average visual eccentricity of round and spiky stimuli (right panel; red and yellow outlines, respectively). (c) Activation of the superior temporal cortex to bouba more than to kiki sounds. This difference was larger in matching than in mismatching bimodal stimuli (right panel; error bars represent ± 1 SEM after subtraction of each subject's global mean; *: $p < 0.05$).

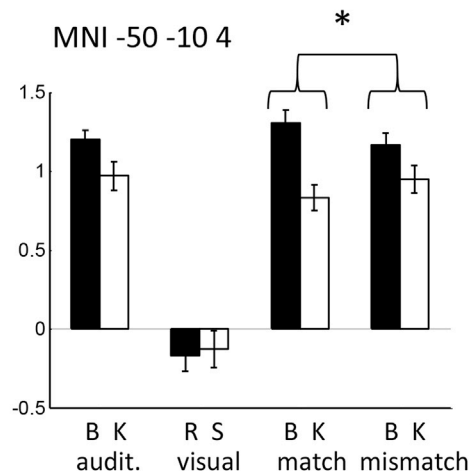
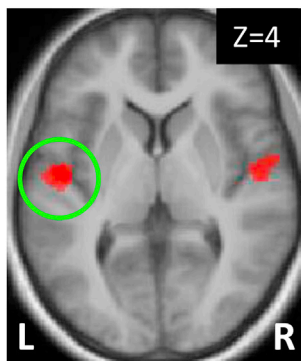
(b) Round vs. spiky shapes



Superimposition of all stimuli



(c) Bouba vs. kiki sounds



Effects of intermodal matching

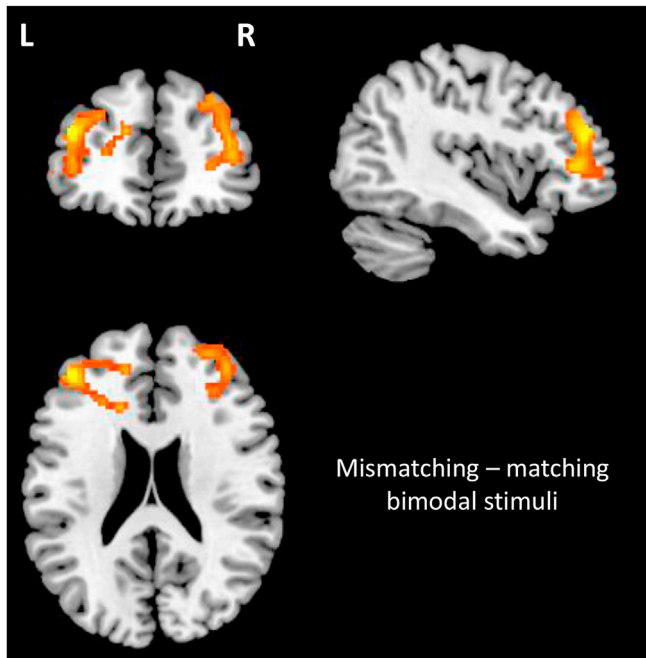


Fig. 4. Bilateral prefrontal activation by mismatching minus matching bimodal stimuli.

difference between bouba and kiki sounds, or between spiky and round shapes, these differences would be modulated by intermodal matching. First, in the left STG/Heschl's region sensitive to the contrast of bouba minus kiki sounds, this difference was larger in matching than in mismatching bimodal trials (at the peak of the bouba minus kiki contrast MNI -50 -10 4; $t(135) = 2.20$; $p = 0.03$) (Fig. 3c, right panel). At the symmetrical right-hemispheric peak, there was a parallel but non-significant trend (MNI 52 -8 2; $t(135) = 1.32$; $p = 0.19$). Second, we studied the occipital regions sensitive to the differences between spiky and round shapes, and did not find modulation of this difference by intermodal matching.

4.3.4. Correlation of intermodal matching with behavior

We looked for regions where the effect of intermodal matching would correlate with individual measures of sensitivity to implicit sound-shape association, as computed from error rates in the previous behavioral experiment. The contrast of matching minus mismatching stimuli was positively correlated with the behavioral congruence score in bilateral occipital and temporal regions activated by shapes relative to rest (Fig. 5a; right: MNI 44 -78 0, $Z = 4.16$; left: MNI -44 -86 -10, $Z = 4.27$, 324 voxels). We repeated the same analysis using an index of sensitivity based on the difference in mean RT (rather than in error rate) between incongruent and congruent blocks. We found no significant activations at the usual threshold. However, the two highest subthreshold clusters were located in the left and right ventral occipital regions (MNI -46 -70 -10; $Z = 3.93$; MNI 40 -78 -18; $Z = 3.66$), precisely overlapping with the significant clusters derived from error rates, suggestive of congruent but noisier correlations.

To better understand the link between activation and behavior, we computed the individual value of this contrast averaged across voxels within those group-level clusters, and plotted it against the behavioral BK effect (Fig. 5b). The correlation pattern of the two variables was as follows. First, the BOLD matching effect was positive in 5 participants and negative in 11, and did not on average differ from zero (average -0.03 ;

Correlation of intermodal matching with behavior

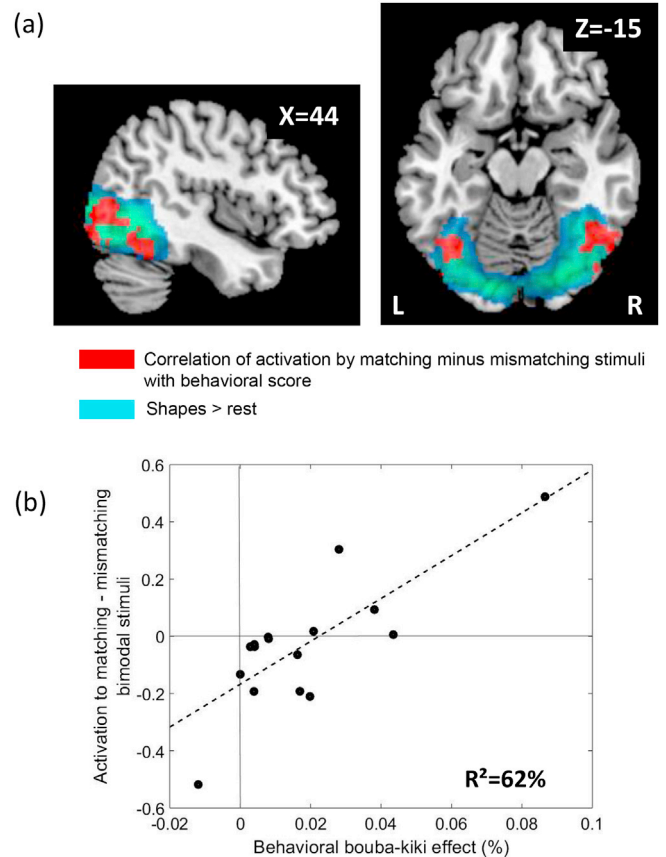


Fig. 5. (a) The activation contrast of matching minus mismatching stimuli was positively correlated with the behavioral congruence score in bilateral occipital and temporal regions (red), which were also activated by shapes relative to rest (green). (b) Plot of individual values of activation and of congruence score.

$p = 0.57$). This is why the matching minus mismatching contrast did not activate the visual cortex at the whole-group level. Second, the behavioral BK effect was positive in 15/16 participants, and was significant on average ($p = 0.0072$), as reported before in the behavioral study. Third, there was a linear increase of the BK effect proportional to the BOLD matching effect ($r = 0.79$; $p = 0.0003$).

Finally, we returned to the previous analysis, computed the individual value of the effect of matching on the bouba-kiki difference in the auditory cortex (MNI -50 -10 4), and correlated it with the behavioral index. This effect was positive in 14/16 participants, and was not correlated with individual behavioral sensitivity to the BK effect ($r = -0.28$; $p = 0.3$).

4.4. Summary

Activations relevant to the implicit BK effect may be summarized as follows.

First, there was a significant difference between mismatching minus matching bimodal stimuli in the bilateral prefrontal cortex (PFC). Second, in the left STG/Heschl's gyrus, the difference between bouba and kiki sounds was larger in matching than in mismatching trials. Third, the difference in activation for matching minus mismatching trials was positively correlated with the behavioral sensitivity to intermodal congruence in the occipitotemporal visual cortex.

5. Discussion

5.1. Optimizing the BK effect

In the first experiment, we successfully replicated the BK effect with a classical paradigm, explicitly asking participants to judge the correspondence between abstract shapes and pseudowords. We found that both vowels and consonants had a correspondence with shape, and that different phonetic features (vowel backness and rounding, consonant voicing and mode of articulation) yielded additive effects (D'Onofrio, 2014). Importantly, the pattern of results was highly consistent across items, indicating that the BK effect prevails with a variety of stimuli much beyond the original and often used stimuli designed by Köhler (1929). We also found that visual size was associated only to vowel type, but not to consonant type. Indeed, small/large visual size has a well-documented association with low/high auditory pitch (Gallace and Spence, 2006; Parise and Spence, 2012), but also with the [i]/[a] vocalic contrast (Parise and Spence, 2012) and with vowel formants (Knoeferle et al., 2017). Finally, the relative strength of the shape-vowel and shape-consonant association varied substantially across participants. Eventually, in order to maximize the BK effect in subsequent experiments, we decided to manipulate shape and not size, and to combine rather than cross the effects of vowel type and consonant type.

5.2. Automaticity of the BK effect

In a careful discussion of automaticity in the context of crossmodal correspondences, Spence and Deroy (2013) put forward the four criteria of goal-independence, non-consciousness, load-insensitivity, and speed. The classical method for assessing the BK effect obviously falls short of those criteria. The interaction of modalities could occur only at a late stage, at which participants would consciously bring to bear semantic knowledge, and bias their decision accordingly, whenever they judge it relevant to the current task. In our second experiment, we tried to determine whether the BK effect would persist even when crossmodal correspondences are task-irrelevant, and participants are presumably unaware of the topic of the study. To this end, we resorted to a method derived from the Implicit Association Test paradigm (Greenwald et al., 1998). A similar approach was adopted by Parise and Spence (2012) for the study of various types of crossmodal correspondences. Among these, they tested the BK effect, albeit only using the two original shapes and pseudowords proposed by Köhler (1929), and found better performance in congruent than in incongruent blocks. Lacey et al. (2016), also using two shapes and two pseudowords, found a significant BK effect in synaesthetes but not in control participants. We extended these results using a broader set of stimuli, and showed that this finding is confirmed beyond the specific set of the original Köhler stimuli. The persistence of the BK effect even in this setting suggests that it may come at least in part from automatic perceptual stages of stimulus encoding, immune from attention and task-related influences. In a recent study, Getz and Kubovy (2018) studied crossmodal correspondences between pitch and various visual features, while jointly manipulating congruence and task instructions. They concluded that even seemingly automatic correspondences always included a top-down component. A further step in the investigation of the automaticity of the BK effect should consist in determining whether sound-shape integration occurs with fully subliminal stimuli, as may happen with audiovisual speech (Faivre et al., 2014) (for reviews on consciousness and multisensory integration, see Deroy et al., 2014; Mudrik et al., 2014).

5.3. The BK effect and object categorization

We make sense of the infinite variety of perceptual experiences by reducing them to a limited set of manageable categories (Seger and Miller, 2010). Categorization may be based on input in one sensory modality, but it generally takes advantage of any source of multimodal

information, based on the ‘unity assumption’, i.e. the assumption that multiple unisensory cues result from one and the same event (Chen and Spence, 2017). Categorization is then easier when all sources of information are congruent, i.e. converge on the same category. For instance, converging on a category may be challenging when facing an animal with the visual features of a dog but emitting the mewing of a cat (Hein et al., 2007). This is also true for simpler perceptual decisions affected by crossmodal correspondences: deciding that an image is “up” is more difficult when primed by an incongruent low-pitch sound than by a high-pitch sound (Evans and Treisman, 2010). Similarly, the congruence between high/low auditory pitch and small/large visual images shown at different locations increases the error in sound localization, known as the ventriloquist effect (Bien et al., 2012). In a Bayesian framework, the unity assumption may be modeled as a coupling prior, that is to say the probability, prior to the present evidence, that the two unisensory signals have a common cause rather than two different causes. Then the advantage of congruent over incongruent stimuli, or binding tendency, reflects the value of the coupling prior for a given individual and a given task (Odegaard and Shams, 2016).

Correlates of such congruence effects have been found in a broad variety of brain regions, depending on task requirements, modality, type of stimuli, sensory or semantic features relevant to categorization, etc (Ghazanfar and Schroeder, 2006). These regions include both higher order association cortex, and unisensory areas. Considering the present results on the correlates of the BK effect, we will first discuss the role of the prefrontal cortex, and then of the auditory and visual regions.

5.4. Involvement of the prefrontal cortex in the BK effect

Involvement of the prefrontal cortex has been demonstrated in a broad variety of categorization processes, in perceptual or abstract rule-based tasks, in both monkeys and humans (for a review, see Seger and Miller, 2010). For instance, in monkeys trained to categorize morphed pictures along a continuum from cats to dogs, lateral prefrontal neurons show sharp differences in activity coinciding with category boundaries (Freedman et al., 2003). Similarly, in human participants trained at discriminating categories along continuous visual dimensions, it is possible to decode categorizing decisions based on the pattern of prefrontal activation (Li et al., 2009).

Prefrontal regions are activated whenever categorization depends on the integration of multisensory cues. As a rule, prefrontal activations are then more intense for incongruent than for congruent stimuli, in agreement with the present finding of stronger activation for mismatching than for matching bimodal stimuli. This phenomenon has been observed in non-linguistic paradigms. Hein et al. (2007) found evidence of audiovisual integration in the IFG, with higher activations for incongruent vs congruent pictures and sounds, a finding which extends to newly learned multimodal objects (Naumer et al., 2009). More recently, McCormick et al. (2018) showed that the congruency of auditory pitch and visual elevation modulated activations in the bilateral prefrontal cortex.

Prefrontal sensitivity to congruence has also been shown in the field of communication and speech processing. In rhesus monkeys, Diehl and Romanski (2014) found that ventrolateral prefrontal neurons show a significant change in neuronal activity in response to movies depicting incongruent versus congruent faces and vocalizations. In humans, Noppeney et al. (2008) found larger prefrontal activations during the categorization of words or sounds, when they were preceded by incongruent than by congruent visual cues. In a related paradigm, Noppeney et al. (2010) showed that the prefrontal incongruence effect was mostly present when targets were perceptually degraded and unreliable, pointing to a role in the accumulation of audiovisual evidence. Moreover, video clips of a speaker uttering incongruent visual and auditory speech yield larger activation of the left IFG than congruent stimuli (Nath and Beauchamp, 2012; Ojanen et al., 2005).

Thus, the higher prefrontal activation which we found for mismatching over matching stimuli fits within a broader pattern observed

repeatedly in crossmodal paradigms. The precise interpretation of those prefrontal activations is however unclear. On the basis of their localization, one may assume that they correspond to decisional processes related to crossmodal integration. This is not in contradiction with our claim that the BK phenomenon was implicit and task-irrelevant in the fMRI paradigm. One possible account is that the visual and auditory BK stimuli tended to distract participants from their explicit task of detecting beeps or crosses, requiring some executive effort, subtended by prefrontal activation, to keep attention focused on the relevant targets (Bisley, 2011; Noudoost et al., 2010). Matching stimuli are automatically and effortlessly merged into one single perceptual object, while mismatching stimuli keep on being represented as distinct events. During mismatching blocks, those distinct events would yield stronger distraction from the task, inducing higher executive effort, and correlatively stronger prefrontal activations.

5.5. Involvement of sensory cortices in the BK effect

Beyond the prefrontal cortex, many of the studies reviewed before also find an impact of crossmodal congruency on sensory cortices. This is in line with the present findings of modulation of auditory areas by audiovisual matching (Fig. 3), and of visual areas by individual sensitivity to the BK effect (Fig. 4). Thus Freedman et al. (2003) found that category effects affected fewer neurons and were more restricted in time, in inferotemporal than in prefrontal cortex. They suggest that the diagnostic features of stimuli may be emphasized in inferotemporal cortex, while the combination of features into an explicit category representation may rely more on prefrontal regions. In humans, Li et al. (2009) found that high-level occipitotemporal areas also show category learning, but they are not sensitive to the current task-requirements, while prefrontal regions are. Such findings support the natural idea of respectively more perceptual and more cognitive contributions of sensory and prefrontal cortices to categorization.

This general view extends to cases of multisensory integration, in which congruence effects also emerge in sensory cortices. Thus, increased activation was found for incongruent vs congruent audiovisual stimuli in the right STS and bilateral STG (Hein et al., 2007) and in the right fusiform region (Noppeney et al., 2010). Concerning auditory speech perception, Noppeney et al. (2008) found increased activation in the left MTG/STS for word targets preceded by incongruent visual cues. The left STS also shows preference for congruent vs incongruent audiovisual syllables (Nath and Beauchamp, 2012). Van Atteveldt et al. (2004, 2006) also found stronger activation for congruent than incongruent printed/auditory vowels in the bilateral superior temporal gyrus.

While we found effects of multisensory congruency in both auditory and visual cortices, these effects appeared in different types of analyses, which may result from some underlying asymmetry in the formally symmetrical experimental design. On the one hand, in the auditory cortex, we found the difference between bouba vs kiki sounds to be larger in matching than in mismatching stimuli, revealing an influence of visual cues on the auditory encoding of speech in the superior temporal cortex. This effect was highly reproducible across participants, while the effect of matching in the visual cortex concerned only a subset of participants, and appeared in a correlation analysis. Indeed, speech and vision entertain asymmetric links in the context of audiovisual speech. Thus, the McGurk effect, a paradigmatic instance of speech-related audiovisual integration, consists in a one-way influence of visual information on the perception and categorization of auditory syllables. Accordingly, the neural signature of the McGurk effect has been identified in supramodal cortex, in superior temporal auditory cortex, but not in visual areas (for a review, see Alsius et al., 2018). More generally, visual cues may be considered as providing automatic support for the dominant auditory speech processing. In the current study, the amplification of the bouba-kiki difference in the auditory cortex may reflect an increased gain of auditory processing when supported by congruent visual information, present in almost all participants. On the other hand, in the visual cortex, we found a positive

correlation of the matching minus mismatching difference with individual sensitivity to the BK effect, reflecting an influence of auditory cues on the visual encoding of shapes in ventral occipitotemporal regions. Due to the variability of this matching effect across participants, this region did not appear in the matching contrast in the main model. In accordance with the non-dominant role of vision in speech perception, an impact of audition on the visual cortex appears to be variable and present only in a subset of participants.

6. Conclusion

In a series of experiments devoted to the BK effect, we addressed two issues. First, using an IAT paradigm, we have shown that the core phenomenon prevails even when the link between round/spiky shapes and bouba/kiki speech sounds is implicit and irrelevant to the task. This suggests that the audiovisual correspondence underlying the BK effect does not require task-dependent effortful decisions, and stems at least in part from an early sensory origin. Second, looking for the brain correlates of implicit BK matching, we found that, in accordance with the previous behavioral finding, intermodal matching influenced activations in auditory and visual sensory cortices. Moreover, we found higher prefrontal activation to mismatching than to matching stimuli, reflecting a modulation by crossmodal correspondence of executive processes. Thus, through its roots in the physiology of object categorization and crossmodal matching, the BK effect provides a unique insight into some non-linguistic components of word formation.

Acknowledgements

This work was supported by the “Investissements d’avenir” program of the French National Research Agency (ANR-10-IAIHU-06; to the Brain and Spine Institute).

References

- D’Onofrio, A., 2014. Phonetic detail and dimensionality in sound-shape correspondences: refining the bouba-kiki paradigm. *Lang. Speech* 57, 367–393.
- Alais, D., Newell, F.N., Mamassian, P., 2010. Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* 23, 3–38.
- Alsius, A., Paré, M., Munhall, K.G., 2018. Forty years after Hearing lips and seeing voices: the McGurk effect revisited. *Multisensory Res.* 31, 111–144.
- Bien, N., ten Oever, S., Goebel, R., Sack, A.T., 2012. The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. *Neuroimage* 59, 663–672.
- Bisley, J.W., 2011. The neural basis of visual attention. *J. Physiol.* 589, 49–57.
- Bizley, J.K., Maddox, R.K., Lee, A.K.C., 2016. Defining auditory-visual objects: behavioral tests and physiological mechanisms. *Trends Neurosci.* 39, 74–85.
- 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, 165–172.
- Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebr. Cortex* 11, 1110–1123.
- Chen, Y.C., Spence, C., 2017. Assessing the role of the ‘unity assumption’ on multisensory integration: a review. *Front. Psychol.* 8, 445.
- Chen, Y.C., Huang, P.C., Woods, A., Spence, C., 2016. When “Bouba” equals “Kiki”: cultural commonalities and cultural differences in sound-shape correspondences. *Sci. Rep.* 6, 26681.
- Chen, Y.-C., Huang, P.-C., Woods, A., Spence, C., 2018. I know that “Kiki” is angular: the metacognition underlying sound-shape correspondences. *Psychonomic Bull. Rev.* 1–8.
- Chiou, R., Rich, A.N., 2012. Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception* 41, 339–353.
- Cuskley, C., Simmer, J., Kirby, S., 2017. Phonological and orthographic influences in the bouba-kiki effect. *Psychol. Res.* 81, 119–130.
- Deroy, O., Spence, C., 2013. Weakening the case for ‘weak synaesthesia’: why crossmodal correspondences are not synaesthetic. *Psychonomic Bull. Rev.* 20, 643–664.
- Deroy, O., Chen, Y.C., Spence, C., 2014. Multisensory constraints on awareness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130207.
- Diehl, M.M., Romanski, L.M., 2014. Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J. Neurosci. : Off. J. Soc. Neurosci. Off. J. Soc. Neurosci.* 34, 11233–11243.
- Evans, K.K., Treisman, A., 2010. Natural cross-modal mappings between visual and auditory features. *J. Vis.* 10, 6 1–12.
- Faivre, N., Mudrik, L., Schwartz, N., Koch, C., 2014. Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. *Psychol. Sci.* 25, 2006–2016.

- Fort, M., Martin, A., Peperkamp, S., 2015. Consonants are more important than vowels in the bouba-kiki effect. *Lang. Speech* 58, 247–266.
- Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246.
- Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., Henson, R.N., 2006. A critique of functional localisers. *Neuroimage* 30, 1077–1087.
- Gallace, A., Spence, C., 2006. Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* 68, 1191–1203.
- Getz, L.M., Kubovy, M., 2018. Questioning the automaticity of audiovisual correspondences. *Cognition* 175, 101–108.
- Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? *Trends Cognit. Sci.* 10, 278–285.
- Greenwald, A.G., Banaji, M.R., 2017. The implicit revolution: reconceiving the relation between conscious and unconscious. *Am. Psychol.* 72, 861–871.
- Greenwald, A.G., McGhee, D.E., Schwartz, J.L., 1998. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.
- Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci. : Off. J. Soc. Neurosci. Off. J. Soc. Neurosci.* 27, 7881–7887.
- Kilian-Hütten, N., Formisano, E., Vroomen, J., 2017. Multisensory integration in speech processing: neural mechanisms of cross-modal aftereffects. In: Mody, M. (Ed.), *Neural Mechanisms of Language: Innovations in Cognitive Neuroscience*. Springer, Boston, MA, pp. 105–127.
- Knoeferle, K., Li, J., Maggioni, E., Spence, C., 2017. What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Sci. Rep.* 7, 5562.
- Köhler, W., 1929. *Gestalt Psychology*. Liveright, New York.
- Köhler, W., 1947. *Gestalt Psychology: an Introduction to New Concepts in Modern Psychology*.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535.
- Lacey, S., Martinez, M., McCormick, K., Sathian, K., 2016. Synesthesia strengthens sound-symbolic cross-modal correspondences. *Eur. J. Neurosci.* 44, 2716–2721.
- Li, S., Mayhew, S.D., Kourtzi, Z., 2009. Learning shapes the representation of behavioral choice in the human brain. *Neuron* 62, 441–452.
- Martino, G., Marks, L.E., 2001. Synesthesia: strong and weak. *Curr. Dir. Psychol. Sci.* 10, 61–65.
- Maurer, D., Pathman, T., Mondloch, C.J., 2006. The shape of boubas: sound-shape correspondences in toddlers and adults. *Dev. Sci.* 9, 316–322.
- McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C., Sathian, K., 2018. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia* 112, 19–30.
- Mudrik, L., Favier, N., Koch, C., 2014. Information integration without awareness. *Trends Cognit. Sci.* 18, 488–496.
- Nath, A.R., Beauchamp, M.S., 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage* 59, 781–787.
- Naumer, M.J., Doehrmann, O., Müller, N.G., Muckli, L., Kaiser, J., Hein, G., 2009. Cortical plasticity of audio-visual object representations. *Cerebr. Cortex* 19, 1641–1653.
- Noppeney, U., Josephs, O., Hocking, J., Price, C.J., Friston, K.J., 2008. The effect of prior visual information on recognition of speech and sounds. *Cerebr. Cortex* 18, 598–609.
- Noppeney, U., Ostwald, D., Werner, S., 2010. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J. Neurosci.* 30, 7434–7446.
- Noudoost, B., Chang, M.H., Steinmetz, N.A., Moore, T., 2010. Top-down control of visual attention. *Curr. Opin. Neurobiol.* 20, 183–190.
- Odegaard, B., Shams, L., 2016. The brain's tendency to bind audiovisual signals is stable but not general. *Psychol. Sci.* 27, 583–591.
- Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I.P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *Neuroimage* 25, 333–338.
- Palmer, S.E., Schloss, K.B., Xu, Z., Prado-León, L.R., 2013. Music-color associations are mediated by emotion. *Proc. Natl. Acad. Sci.* 110, 8836–8841.
- Parise, C.V., Spence, C., 2012. Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Exp. Brain Res.* 220, 319–333.
- Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., Banaji, M.R., 2000. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cognit. Neurosci.* 12, 729–738.
- Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E., 2008. Guidelines for reporting an fMRI study. *Neuroimage* 40, 409–414.
- Ramachandran, V.S., Hubbard, E.M., 2001. Synaesthesia—a window into perception, thought and language. *J. Conscious. Stud.* 8, 3–34.
- Revell, K.P., Namy, L.L., DeFife, L.C., Nygaard, L.C., 2014. Cross-linguistic sound symbolism and crossmodal correspondence: evidence from fMRI and DTI. *Brain Lang.* 128, 18–24.
- Sadaghiani, S., Maier, J.X., Noppeney, U., 2009. Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J. Neurosci. : Off. J. Soc. Neurosci.* 29, 6490–6499.
- Sawilowsky, S.S., 2009. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* 8 (2), 597–599.
- Seger, C.A., Miller, E.K., 2010. Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219.
- Spence, C., 2011. Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995.
- Spence, C., Deroy, O., 2013. How automatic are crossmodal correspondences? *Conscious. Cognit.* 22, 245–260.
- Taitz, A., Assaneo, M.F., Elisei, N., Tripodi, M., Cohen, L., Sitt, J.D., Trevisan, M.A., 2018. The audiovisual structure of onomatopoeias: an intrusion of real-world physics in lexical creation. *PLoS One* 13, e0193466.
- Tarte, R.D., Barritt, L.S., 1971. Phonetic symbolism in adult native speakers of English: three studies. *Lang. Speech* 14, 158–168.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282.
- van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2006. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebr. Cortex* 17 (4), 962–974.
- van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1181–1186.
- Xu, K., Nosek, B., Greenwald, A., 2014. Psychology data from the race implicit association test on the project implicit demo website. *J. Open Psychol. Data* 2.