

ORIGINAL RESEARCH REPORT

How Iconicity Helps People Learn New Words: Neural Correlates and Individual Differences in Sound-Symbolic Bootstrapping

Gwilym Lockwood*, Peter Hagoort*[†] and Mark Dingemans[‡]

Sound symbolism is increasingly understood as involving iconicity, or perceptual analogies and cross-modal correspondences between form and meaning, but the search for its functional and neural correlates is ongoing. Here we study how people learn sound-symbolic words, using behavioural, electrophysiological and individual difference measures. Dutch participants learned Japanese ideophones—lexical sound-symbolic words—with a translation of either the real meaning (in which form and meaning show cross-modal correspondences) or the opposite meaning (in which form and meaning show cross-modal clashes). Participants were significantly better at identifying the words they learned in the real condition, correctly remembering the real word pairing 86.7% of the time, but the opposite word pairing only 71.3% of the time. Analysing event-related potentials (ERPs) during the test round showed that ideophones in the real condition elicited a greater P3 component and late positive complex than ideophones in the opposite condition. In a subsequent forced choice task, participants were asked to guess the real translation from two alternatives. They did this with 73.0% accuracy, well above chance level even for words they had encountered in the opposite condition, showing that people are generally sensitive to the sound-symbolic cues in ideophones. Individual difference measures showed that the ERP effect in the test round of the learning task was greater for participants who were more sensitive to sound symbolism in the forced choice task. The main driver of the difference was a lower amplitude of the P3 component in response to ideophones in the opposite condition, suggesting that people who are more sensitive to sound symbolism may have more difficulty to suppress conflicting cross-modal information. The findings provide new evidence that cross-modal correspondences between sound and meaning facilitate word learning, while cross-modal clashes make word learning harder, especially for people who are more sensitive to sound symbolism.

Keywords: sound symbolism; ideophones; language; ERP; iconicity; individual differences

Introduction

Iconicity, or the resemblance-based mapping between aspects of form and meaning, has long been marginalised in linguistic research due to the predominance of arbitrariness, where there is no connection between the form of a word and aspects of its meaning other than social convention [1]. For example, there is nothing iconic about the arbitrary word *dog*; *d* does not mean “four-legged”, *o* does not mean “pet”, and *g* does not mean “likes rolling in muddy puddles”. Nothing about the form of the word represents the real world meaning. In contrast, the Siwu words *pim-bilii* (‘small belly’) and *pumbuluu* (‘enormous round belly’)

use the vowel space to iconically depict the size of the referent’s belly. In British Sign Language, the sign for ‘tree’ is also iconic, as it features the primary forearm raised, representing the trunk, with the hand open and the fingers spread, representing the branches and leaves [2].

Recent psychological research has shown that iconicity plays a bigger role in language than traditionally thought [3, 4], and that people are sensitive to sound symbolism (which is iconicity specifically for spoken languages) in psycholinguistic tasks. However, much of this research is based on two-alternative forced choice paradigms with pseudowords which are deliberately constructed to maximise iconic contrasts. For a more detailed picture of how sound symbolism works in natural language, pseudoword experiments will need to be supplemented with work using real sound-symbolic words. In spoken languages, one source of sound-symbolic words is the lexical class of ideophones, which are marked words which depict sensory imagery [5, 6].

Behavioural experiments with ideophones have mostly tended to show that people can guess the meanings of ideophones at above chance levels, and that this is

* Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

[†] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, the Netherlands

[‡] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Corresponding author: Gwilym Lockwood (gwilym.lockwood@mpi.nl)

modulated by articulation and prosody. Oda [7] showed that English speakers could guess at above chance levels which Japanese ideophones and English translations went together when hearing the words pronounced by a Japanese speaker, and that their accuracy improved when they articulated the ideophones themselves. Iwasaki et al. [8] also showed that English speakers were sensitive to the meanings of Japanese ideophones, and that English speakers' judgements on the semantic dimensions of the event depicted by an ideophone were broadly consistent with those of Japanese speakers. Meanwhile, Kunihiro [9] found that English speakers could guess the meanings of apparently arbitrary Japanese words better than chance when hearing the words in a monotone voice, and better still when hearing the words in an expressive voice. This shows that even arbitrary words may have residual levels of sound symbolism in them, and that prosody is an important factor in the perception sound symbolism.

Nygaard et al. [10] followed up Kunihiro's study with a word learning experiment. They found that English participants were faster and more accurate at remembering Japanese words taught with their actual English translation than a random English translation, and that there was no difference when learning words with their opposite translations. Nygaard et al. argue that "the sound structure of spoken language may engage cross-modal perceptual-motor correspondences that permeate the form, structure, and meaning of linguistic communication" and that the learners in their experiment were unconsciously able to "exploit non-arbitrary relationships in the service of word learning and retrieval". While we are sympathetic to Nygaard et al.'s arguments, a limitation is that using a variety of nouns, adjectives, and verbs with a variety of prosodic contours and morphophonological structures obscures the many potential sources of sound symbolism that their participants may have identified. Moreover, some of the stimuli used lend themselves more obviously to real and opposite conditions (such as using "slow" as the opposite of *hayai*, which means "fast") than others (such as using "gold" as the opposite of *tetsu*, which means "iron"). Finally, while the words that Kunihiro and Nygaard et al. used in their experiments (and many words in many languages in general) may contain a certain degree of sound-symbolic mappings, they are generally considered to be arbitrary.

In Lockwood, Dingemans, and Hagoort [11], we ran a learning experiment similar to the Kunihiro and Nygaard studies, but strictly with Japanese ideophones, which were controlled for length, grammatical category, and morphophonological structure, and which are strongly sound-symbolic. We showed that Dutch adults learned novel Japanese ideophones better when they were learned with their real Dutch translations (i.e. when there was a sound-symbolic relationship between form and meaning) than when they were learned with their opposite Dutch translations (i.e. when there was either no match or a mismatch between form and meaning). We then informed participants of the manipulation, and asked them to choose what they thought the best translation would be in a two-alternative forced choice task. Despite the learning task, the participants were still sensitive to the

ideophones' meanings and guessed well above chance at 72% accuracy. Meanwhile, we ran the same manipulation with a set of arbitrary adjectives –i.e. adjectives which are not ideophones and are not considered sound-symbolic– with a second group of participants. Participants were able to guess the meanings of the words in a two-alternative forced choice test above chance at 63% accuracy, but the learning effect disappeared completely; participants remembered the adjectives with their real translations at the same level of accuracy as the adjectives with their opposite translations, echoing Nygaard et al.'s [10] findings. We used Japanese for consistency with earlier studies and because Japanese is probably the most well-documented language with an extensive set of ideophones [12, 13], but given the typological unity of ideophones [14, 15] we are fairly confident that this effect would hold in any language with ideophones.

There has been relatively little neuroimaging work on sound symbolism involving real words. In a sentence reading experiment with native Japanese speakers, Lockwood & Tuomainen [16] found that ideophones elicit a greater P2 component and elicit a larger late positive complex (LPC) compared to arbitrary words. They argue that the P2 reflects the multisensory integration of sounds and the associated sensory representations, and that the LPC may reflect higher processing demands of ideophones. In fMRI experiments, Revill et al. [17] and Kanero et al. [18] have both found that sound-symbolic words activate certain brain areas more strongly than non-sound-symbolic words. Revill et al. used words from a variety of languages (some of them historically related) and from a variety of word classes, and labelled the words which English speakers were better able to guess the meanings of as "sound-symbolic", while words that they guessed at chance were labelled as "non-sound-symbolic". The sound-symbolic words elicited more activation than the non-sound-symbolic words in intraparietal areas associated with cross-modal and synaesthetic processing. Kanero et al. compared Japanese ideophones with arbitrary words when participants viewed matching or mismatching videos of motions and images of shapes. They found that the ideophones uniquely activated the right posterior superior temporal sulcus, and that this activation was greater when the ideophones and the videos/images were rated as better matching. They speculate that the right posterior STS integrates the processing of linguistic and environmental sounds. There is very little event-related potential (ERP) research on sound symbolism in real words, and Kanero et al. and Revill et al. make similar arguments about different brain areas in fMRI research. This means that more neuroimaging work is needed in order to work out how the brain processes sound symbolism.

Here we build on this work, extending it in two ways to advance our understanding of sound symbolism. First, we used ideophones, words considered strongly sound-symbolic or iconic by both linguists [14, 19] and native speakers [20, 21]. Using a more unified and linguistically and prosodically homogeneous set of words makes it easier to eliminate possible confounds and be confident that any effect we find is a reliable indicator of sound symbolism.

We repeated the behavioural task in Lockwood et al. [11] with minor alterations to measure the participants' brain activity using EEG (electroencephalography). As Lockwood et al. showed that there was no learning effect with regular arbitrary adjectives, only ideophones were used in the current study.

Second, we analysed event-related potentials (ERPs) to explore the neural mechanisms underlying the processing of sound-symbolic words. We used ERPs to look at the time course of the neural effect; if an early effect was present, as in Kovic et al.'s [22] study with pseudowords, this would suggest that the effect is based on differences in the processing of the sensory properties of the stimuli, whereas if the effect was much later, it would suggest a more linguistic mechanism. It is possible that there are both sensory and linguistic effects, as suggested in ERP experiments by Lockwood and Tuomainen [16] and Sučević et al. [23].

Coupling behavioural data and brain imaging allows us to investigate possible individual differences in sound-symbolic sensitivity. The topic of individual differences [24] has barely been broached in the sound-symbolism literature so far, but is likely to be of key importance in the quest for causal models of sound-symbolism.

We hypothesised that we would behaviourally replicate Lockwood et al. [11], namely that participants would learn the ideophones in the real condition better than in the opposite condition and that participants would still be sensitive to the meanings of ideophones in the two-alternative forced choice task afterwards despite the learning rounds. We also predicted that there would be a correlation between the reaction time and accuracy of judgement of ideophones, in that the more accurately guessed ideophones would also be more quickly guessed. As for the ERP results, since the few sound symbolism ERP studies so far have found different components, we used a non-parametric cluster-based permutation test to investigate the data before analysing particular windows. Finally, we investigated individual differences in the data by looking at the relation between the ERP effect size, the memory/learning performance of the task, and behavioural measures of sensitivity to sound symbolism per participant. We did this in order to see whether the effect was more related to participants' sensitivity to sound symbolism or more related to participants' general task performance.

Methods

Stimuli

This experiment used the same paradigm as Lockwood et al., [11].

We used 38 Japanese ideophones with a reduplicated CVCV-CVCV pattern (see **Table 1** for examples). Ideophones and Dutch translations were matched for word length and

characters in common across conditions. Dutch translations across conditions were matched for word frequency in the Celex corpus (mean log frequency real: 7.68, mean log frequency opposite: 8.18). Participants learned the real translations to 19 ideophones and the opposite translations to the other 19 ideophones. In a published pretest using a fully counterbalanced set of stimuli [11], we found a main effect of real vs. opposite condition in both groups. As counterbalancing made no difference to the results, the stimuli we use here are consistent across participants; all learned *fuwafuwa* as *pluizig*, for example.

Procedure

Participants were told that they were going to learn 38 Japanese words, and that they had to remember the word pairs for a recognition test straight after the learning rounds. After the test, participants were informed that half the words they had learned were correct, but half were the opposite meaning. We then asked them to ignore what they had just learned and instead choose which translation they felt was more natural for each ideophone during the 2AFC task.

Participants saw each ideophone and translation once in a learning round; there were two learning rounds in total. The order of Dutch words and ideophones was randomised for each round and for each participant. We used Presentation to present stimuli and record responses.

The initial Dutch word was presented for 1000ms with 100ms of jitter each way (i.e. between 900ms and 1100ms), followed by a fixation cross for 1000ms with 100ms of jitter. As the ideophone was played over the speakers, a blank screen was presented for 2000ms with 200ms of jitter. This was again followed by a fixation cross. The final screen with the ideophone and its Dutch meaning was presented until participants were happy to move onto the next item. Between trials, a blank screen was presented for 1000ms with 200ms of jitter, followed by a fixation cross for 1000ms with 100ms of jitter to announce the beginning of the next trial.

When it came to the test round, participants were presented with either the word pairs that they had learned (for example, *fuwafuwa* and *pluizig* in the real condition, and *kibikibi* and *futloos* in the opposite condition), or a pseudo-randomised pairing of ideophones and translations which they had seen before. These pairings were pseudo-randomised to ensure that the meanings were semantically unrelated (for example, the Japanese *fuwafuwa*, learned as “fluffy”, and the Dutch *kortaf*, meaning “curt”). Participants were instructed to indicate whether this was a word pair they had learned by answering Yes (left CTRL key) or No (right CTRL key). Pairs requiring a Yes response made up 50% of the trials. As in the learning

| REAL condition | | OPPOSITE condition | |
|-----------------------|------------------------|------------------------|-------------------------|
| ideophone | translation | ideophone | translation |
| fuwafuwa (“fluffy”) | pluizig (“fluffy”) | kibikibi (“energetic”) | futloos (“tame, tired”) |
| boroboro (“worn out”) | versleten (“worn out”) | ukiuki (“happy”) | verdrietig (“sad”) |

Table 1: Example stimuli for each condition.

round, participants saw the Dutch word first, then heard the Japanese ideophone for 2000ms. Then, instead of seeing a fixation cross, they saw a question mark. Participants were asked to respond as soon as possible after seeing the question mark.

Timings in the test stage were identical to the learning stage. The question mark was displayed until participants responded, at which point a blank screen was presented, followed by a fixation cross to announce the beginning of the next trial. In order to ensure enough trials for ERP analysis, the test stage was twice as long as in Lockwood et al. [11], so that there were 38 trials per condition (i.e. 19 ideophones with their real translation, 19 ideophones with their opposite translation, and 38 ideophones with a pseudo-randomised wrong translation, all repeated).

After the test round, we implemented a two-alternative forced choice task as a separate measure of sound-symbolic sensitivity. This was to see if, despite the learning phase, participants were still able to make decisions based on the sound symbolism of the ideophones. Participants heard the ideophone, and then saw the two possible Dutch translations; they selected the translation by pressing the left CTRL key for the translation on the left and the right CTRL key for the translation on the right. Timings were identical to the learning and test stages.

The full experiment is illustrated in **Figure 1** below.

Participants

We tested 40 native Dutch speaking participants (10m, 30f) aged 18–29 (mean: 21y 7m) with normal or corrected-to-normal vision, recruited from the MPI participant database. All participants had no knowledge of Japanese, and were students at either the Radboud University or the Hogeschool van Arnhem en Nijmegen. Participants gave informed written consent to take part in the experiment. The experiment was approved by the Ethics Committee for Behavioural Research of the Social Sciences

Faculty at Radboud University Nijmegen in compliance with the Declaration of Helsinki. Participants were paid 8 Euro per hour for their participation. Participants were told that data sharing was optional, and all participants explicitly opted in to consenting to their data being shared.

In order to make sure that we were testing ERPs from participants who had learned the words, we discarded five participants who scored under 60% in the test round and could have just been guessing the answers. A further six participants were discarded due to excessive artefacts (affecting more than 25% of trials). This left 29 participants in the final dataset (7m, 22f; 19–28 years old, mean 21y 9m; 24 right-handed, 5 left-handed).

EEG recording

EEG was recorded from 61 active Ag/AgCl electrodes, of which 59 were mounted in a cap (actiCap), referenced to the left mastoid. Two separate electrodes were placed at the left and right mastoids. Blinks were monitored through an electrode on the infraorbital ridge below the left eye. The ground electrode was placed on the forehead. Electrode impedance was kept below 10 k Ω . EEG and EOG recordings were amplified through BrainAmp DC amplifiers with a bandpass filter of 0.016–100 Hz, digitised online with a sampling frequency of 500 Hz, and stored for off-line analysis.

ERP analysis

Automatic artefact rejection in BrainVision Analyzer discarded all segments with activity exceeding $\pm 75 \mu\text{V}$. In six of the 29 participants used for the ERP analysis, between one and four individual electrodes were removed and interpolated due to faulty connections. ERPs were time-locked to the onset of the ideophone recording. Across the 29 participants used for all analyses reported in this paper, 13.1% of ideophone trials were rejected due to artefacts.

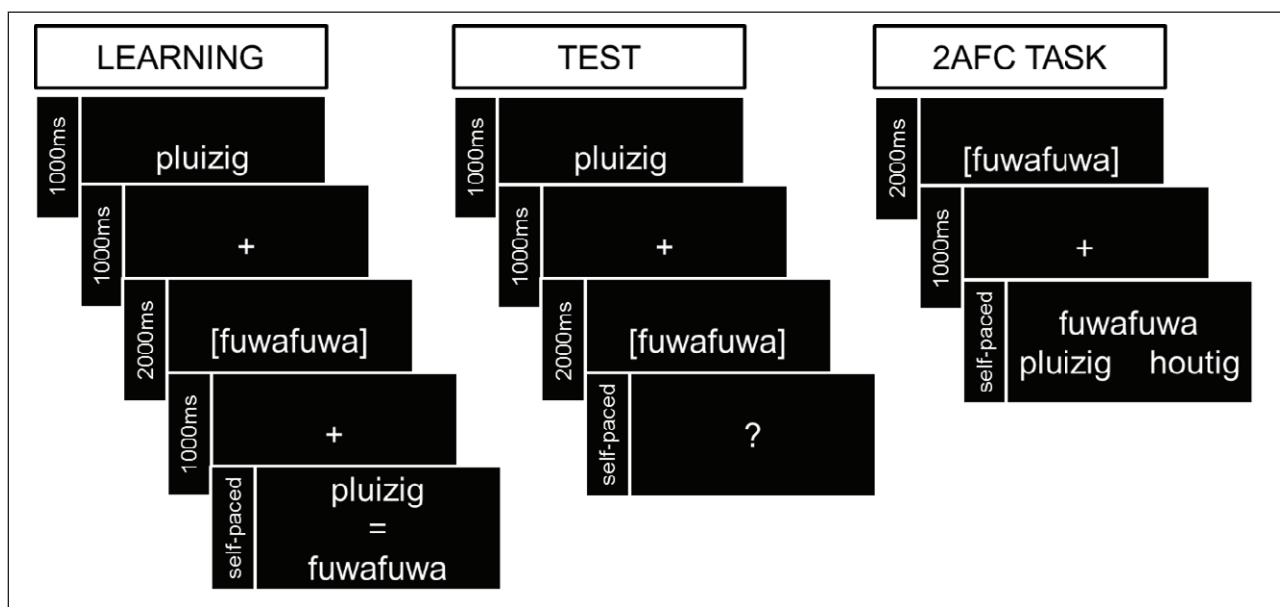


Figure 1: Learning, test, and 2AFC procedure.

As previous sound symbolism studies using ERPs have found mixed results, we used a non-parametric cluster-based permutation test in Fieldtrip [25]. This investigated the entire epoch to establish whether there was a difference between conditions at any given point while correcting for multiple comparisons, and highlighted time windows of interest to analyse. We then ran ANOVAs on mean amplitudes in individual time windows of interest.

Results

Behavioural Results

Main experiment

Participants made more recognition mistakes in the opposite condition than in the real condition; participants correctly remembered the real word pairing 86.7% of the time (95% CIs: 82.92%–90.41%), but correctly remembered the opposite word pairing only 71.3% of the time (95% CIs: 65.19%–77.46%). This is shown in **Figure 2** below, presented in this way rather than as a histogram with error bars in order to better represent the spread of data [26]. Only six out of 29 participants did not show an advantage for the real condition over the opposite condition. Four participants scored higher in the opposite condition than the real condition (with a mean difference of 3.95 percentage points), and two participants had equal scores in both conditions.

As the dependent variable was binary—correct or incorrect—we analysed the responses using a mixed-effects logit

model with the `glmer` function of the `lme4` (versions 1.1-8) package in R. The data was modelled by including a per-participant and per-ideophone random adjustment to the fixed intercept with a random slope for the fixed effect by participant. The condition was sum contrast coded.

Model comparison showed a random effect by ideophone (log likelihood difference = 21.3, $\chi^2 = 42.64$, $df = 1$, $p < 0.001$). That means that some ideophones were answered correctly more often than others. However, even when controlling for this random effect by ideophone, model comparison showed a significant fixed effect of condition ($\beta = -0.5514$, log likelihood difference = 8.2, $\chi^2 = 16.44$, $df = 1$, $p < 0.001$). The model estimated that ideophones learned in the real condition were answered 8.1 percentage points more accurately than ideophones learned in the opposite condition.

There were also significant differences in reaction times between conditions, with participants responding faster to ideophones in the real condition (mean RT = 958ms \pm 95ms CIs) than the opposite condition (mean RT = 1262ms \pm 86ms CIs) ($t = -5.00$, $p < 0.001$, Cohen's $d = -1.63$). This difference existed even when only analysing correctly answered trials ($t = -4.58$, $p < 0.001$, Cohen's $d = -1.49$), and so is not just a speed/accuracy trade off. There was also a strong correlation between the number of correct responses per ideophone and the speed of the reaction to that ideophone; the better an ideophone was remembered, the faster it was responded to ($r = -0.71$,

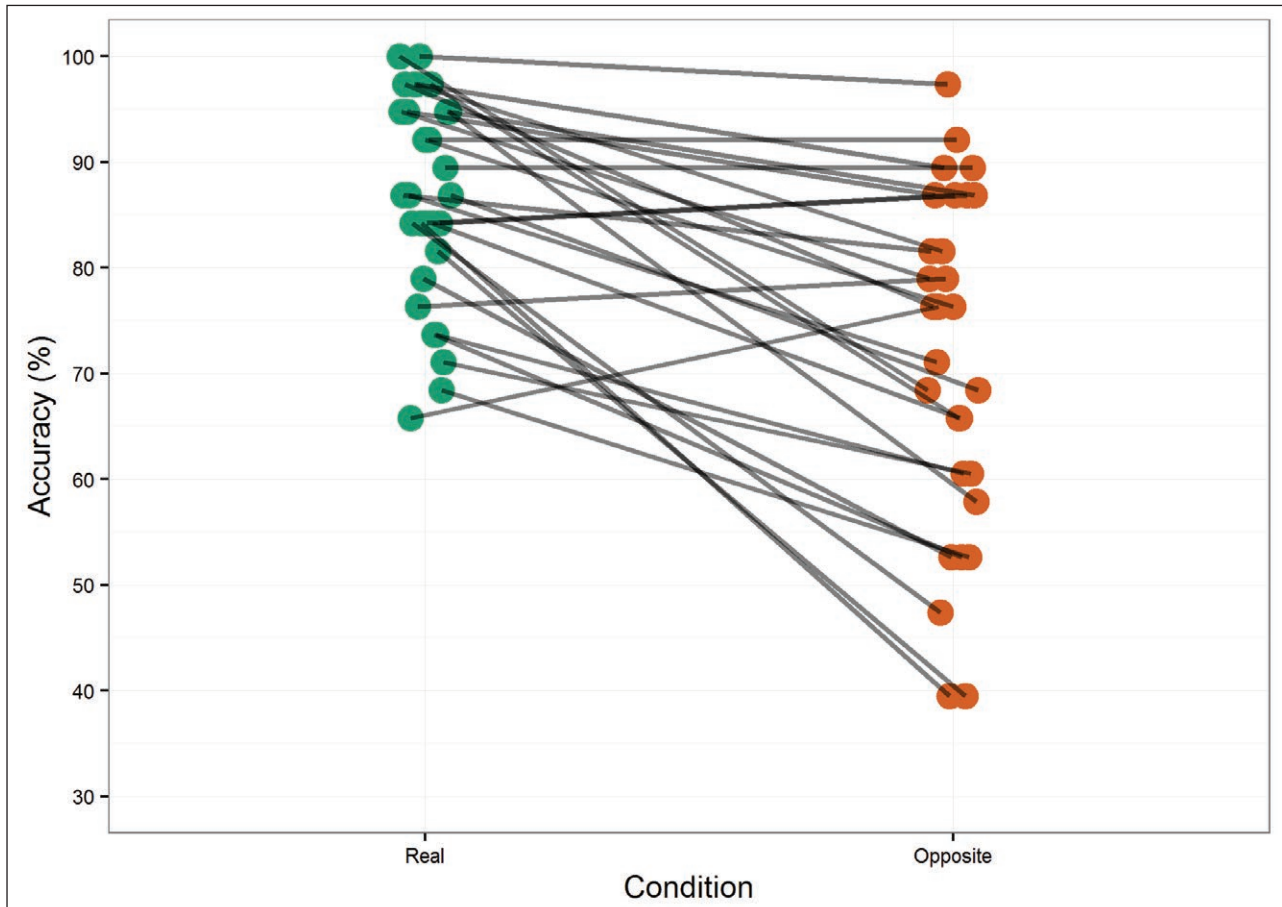


Figure 2: Accuracy per condition (dots represent participants, colours condition).

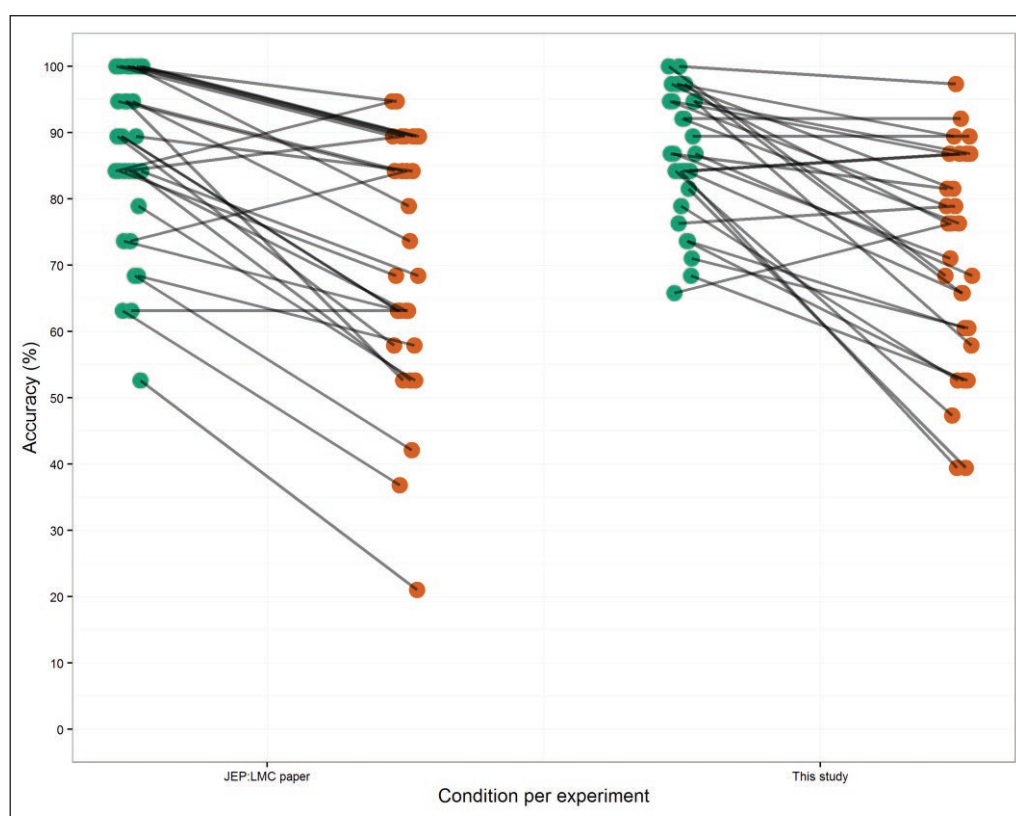


Figure 3: Accuracy per condition (dots represent participants, colours condition) in comparison with our previous behavioural study.

$p < 0.001$). However, there was no correlation between the number of correct responses per participant and reaction times ($r = -0.11$, $p = 0.57$, Cohen's $d = 0.63$), meaning that more accurate participants were not necessarily faster at responding.

This closely replicates the results from Lockwood et al. [11], as shown in **Figure 3**.

Post-experiment sound symbolism sensitivity check

In the sound symbolism sensitivity check after the experiment, participants guessed the real meanings of the Japanese words with 72.96% accuracy, which was comfortably above chance ($\mu = 0.5$, $t = 13.86$, $df = 28$, $p < 0.0001$, 95% CIs = 69.56%–76.35%, Cohen's $d = 5.24$). Only one participant guessed the ideophones at 50% accuracy, and 27 out of 29 participants guessed at least 24 out of 38 ideophones correctly. We checked to see if participants who guessed more accurately also guessed faster, but there was no link between reaction times and accuracy ($r = 0.07$, $n = 29$, $p = 0.73$). Only three ideophones were guessed at below 50% accuracy (*hiyahiya* at 41.4%, *morimori* at 44.8%, *gowagowa* at 48.4%). We also checked to see if ideophones which were guessed more accurately were guessed more quickly. There was a correlation between reaction time and the mean accuracy at which the ideophone was guessed ($r = -0.46$, $n = 38$, $p = 0.0037$). This is shown in **Figure 4**.

One might ask whether the two-alternative forced choice test is affected by taking place *after* the learning round, as participants might continue to select the

words they had learned and maybe only change a few decisions. Participants guessed the real meanings of the words they had previously learned in the real condition at 77.3% (95% CIs: 70.9%–83.7%), and they guessed the real meanings of the words they had previously learned in the opposite condition at 68.6% (95% CIs: 61.9%–75.3%), and this is shown in **Figure 5**. This suggests that participants were still sensitive to sound symbolism, especially as they picked the correct translation of the ideophones originally taught in the opposite condition 68.6% of the time despite being taught explicitly otherwise. However, they may have found it harder to reverse this learning than they did to reevaluate the ideophones they had learned in the real condition; there was a trend towards guessing ideophones previously learned in the real condition more accurately than ideophones previously learned in the opposite condition ($t = 1.9665$, $p = 0.057$, Cohen's $d = 0.64$). This is in line with our predictions in Lockwood et al. [11] that further exposure to ideophones and learned translations decreases the ability to reevaluate sound-symbolic mappings.

Behavioural measures of sound symbolism

Finally, we contrasted the two behavioural measures of sound symbolism in the experiment: the two-alternative forced choice task, and the difference in test scores between the real condition and the opposite condition per participant. Participants who are more sensitive to sound symbolism should find it easier to remember the ideophones in the real condition and harder to remember

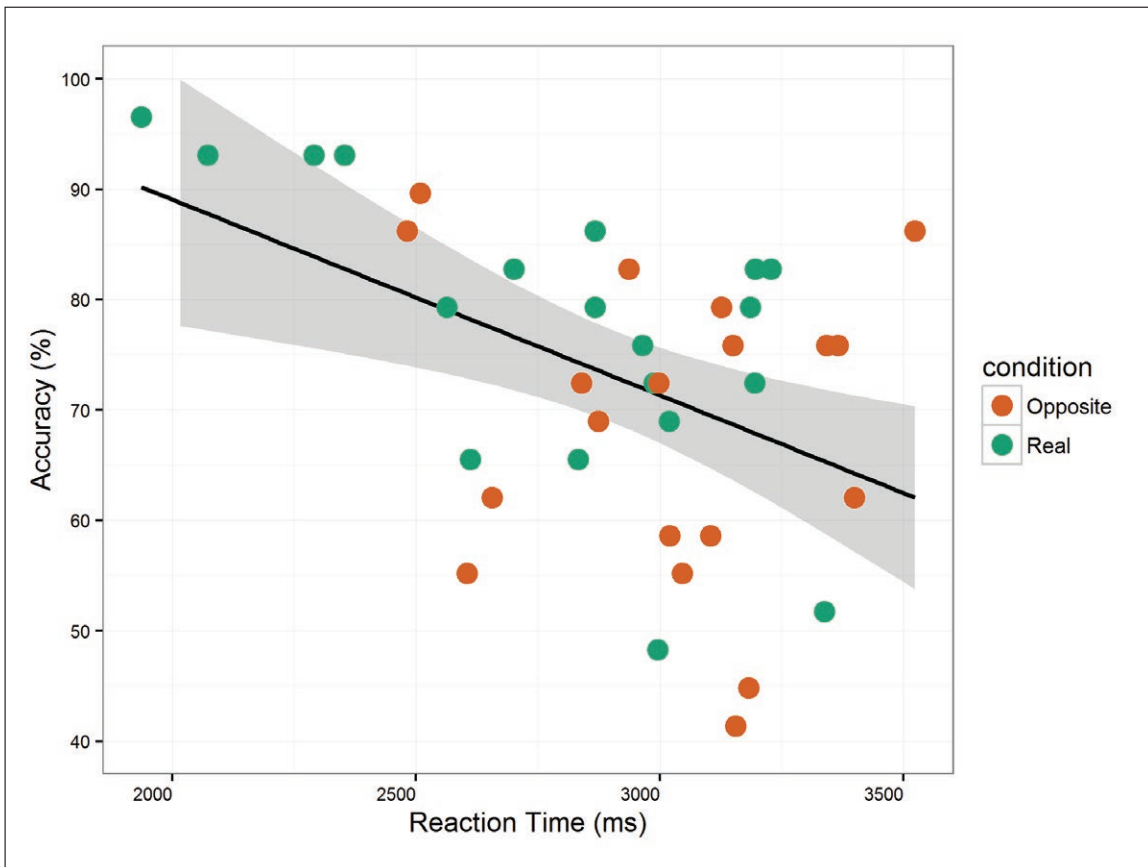


Figure 4: Scatterplot showing correlation between mean accuracy per ideophone and reaction time.

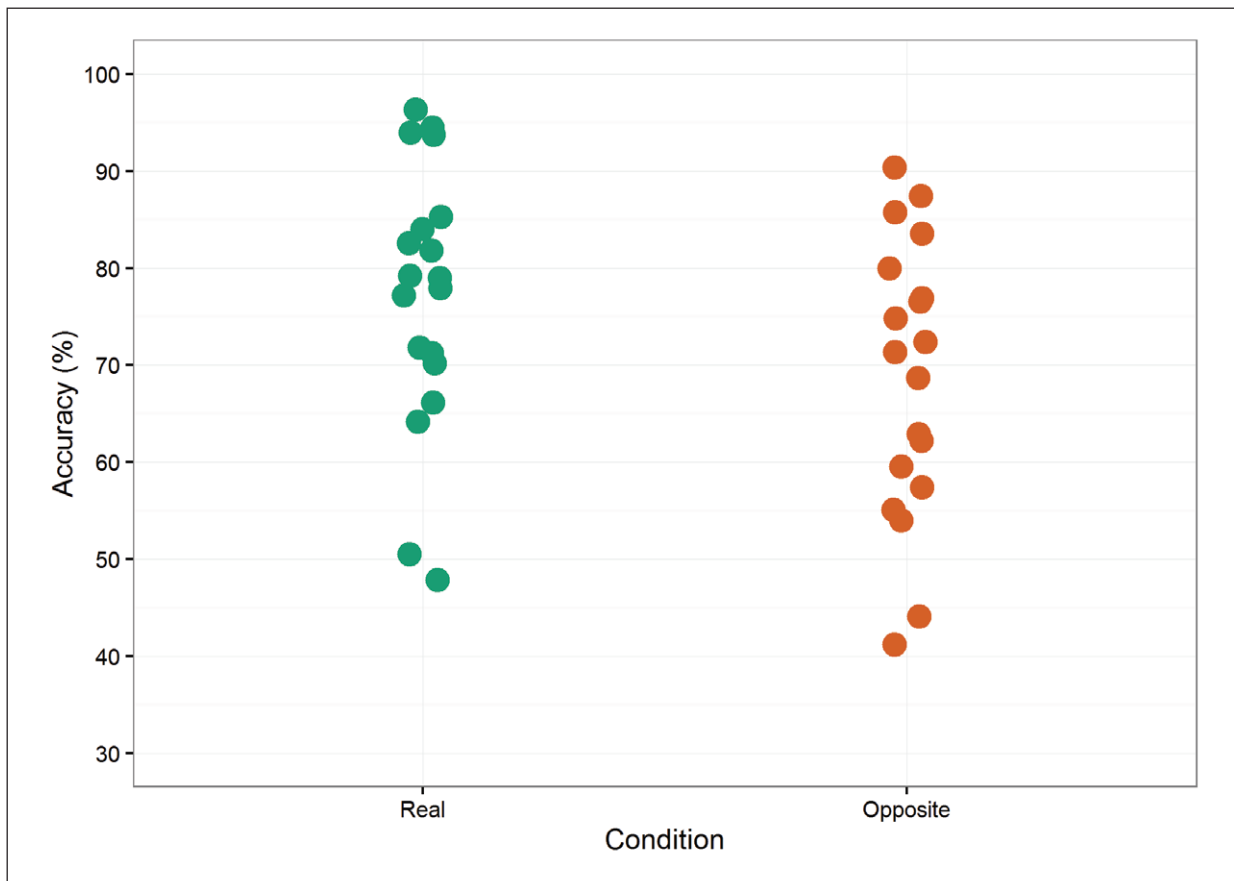


Figure 5: Scatter plot showing the lack of difference in baseline guessing accuracy depending on the condition the ideophone had previously been learned in. Dots represent ideophones.

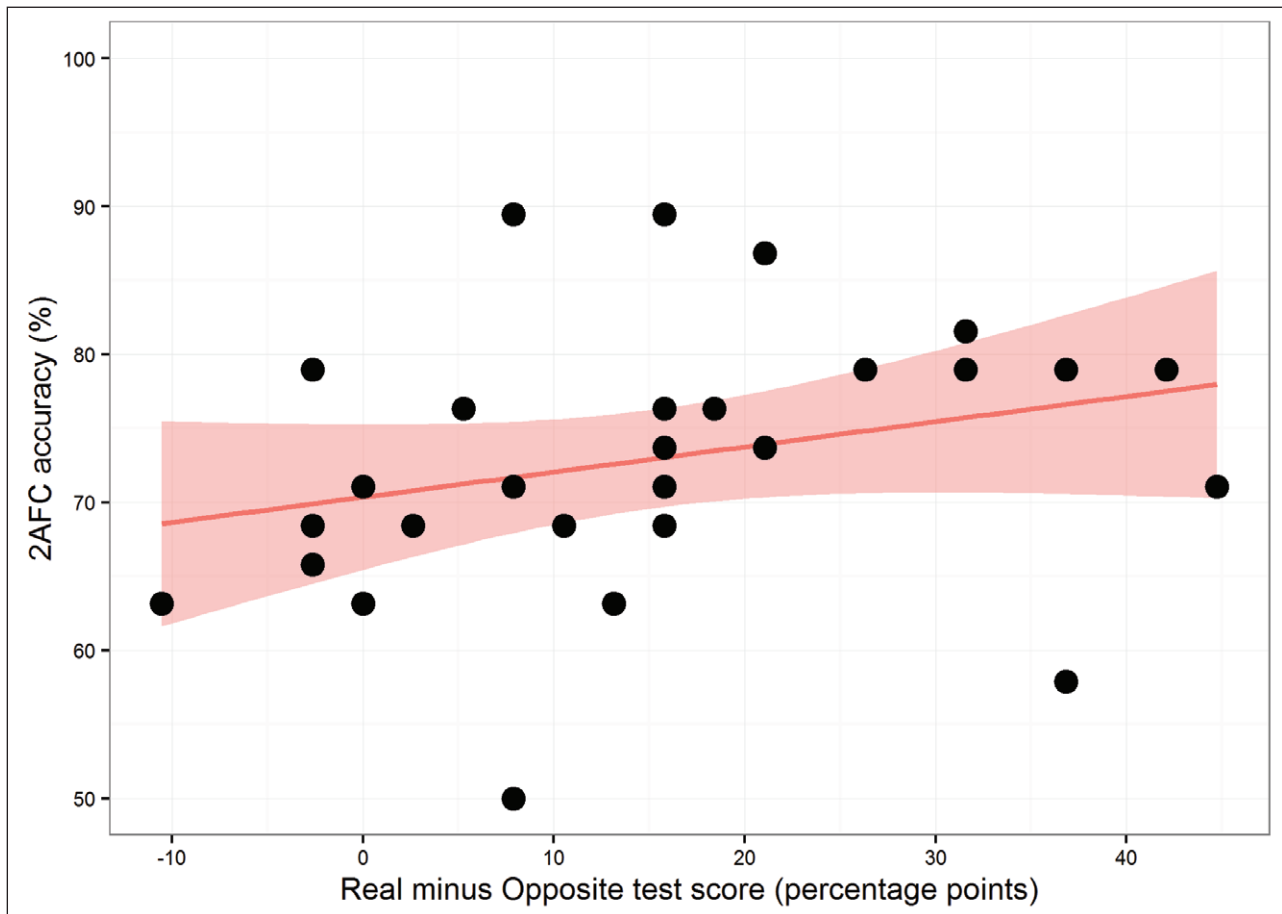


Figure 6: Scatter plot of test score difference and 2AFC task accuracy. Dots represent participants.

the ideophones in the opposite condition; therefore, participants who scored higher in the two-alternative forced choice task should also have a greater disparity in their test scores between conditions.

The two measures were ranked and showed a Spearman correlation ($r = 0.42$, $p = 0.0251$), which suggests that people who are sensitive to sound symbolism when asked to guess a word's meaning are more likely to be affected by that sensitivity during word learning. The correlation is plotted in **Figure 6**.

ERP results

We examined the ERPs from the participants' passive exposure to the ideophones in the learning rounds and from the participants' exposure to the ideophones during the test round.

Somewhat to our surprise, there was no effect of sound symbolism when participants heard the ideophones during the learning rounds. ERPs were time-locked to the onset of the recording of the ideophone in the learning rounds, but there was no effect when looking at the first learning round, the second learning round, or both together. However, there was a considerable effect in the test round.

In the ERPs from the test rounds, we first ran a cluster-based permutation test with 3000 randomisations in Fieldtrip [25] to establish whether there were any differences between real and opposite conditions across the

entire averaged epoch. The cluster-based permutation revealed that there was a significant difference between the two conditions, and that this difference was driven by one cluster starting at 320ms and ending at 786ms ($p = 0.0027$).

Averaged ERP mean amplitudes from nine parietal electrodes (C30, C29, C28, C1, C3, C4, C33, C34, C35) are shown below in **Figure 7**, and topographic plots of the difference between conditions are shown in **Figure 8**. The ERPs are time-locked to the onset of the ideophone. Shading around the ERP lines shows 95% confidence intervals. The topographic plots are calculated by subtracting the opposite condition measurements from the real condition measurements.

We used the cluster and inspection of the waveforms to inform our selection of time windows for further analysis; a P3 effect from 320ms to 500ms, and a late positive complex from 500ms until the end of the cluster at 786ms.

We averaged electrode amplitudes across the midline and four quadrants (left anterior, right anterior, left posterior, right posterior) and ran within-subject 2x5 ANOVAs on the two time windows. In both windows, there was a significant main effect of condition, with ideophones in the real condition eliciting greater a P3 ($F = 16.99$, $df = 1,28$, $p = 0.0003$) and late positive complex ($F = 8.96$, $df = 1,28$, $p = 0.0057$). Interactions between condition and quadrant were not significant for the P3 ($p = 0.051$) or late positive complex ($p = 0.17$). Although the interaction was

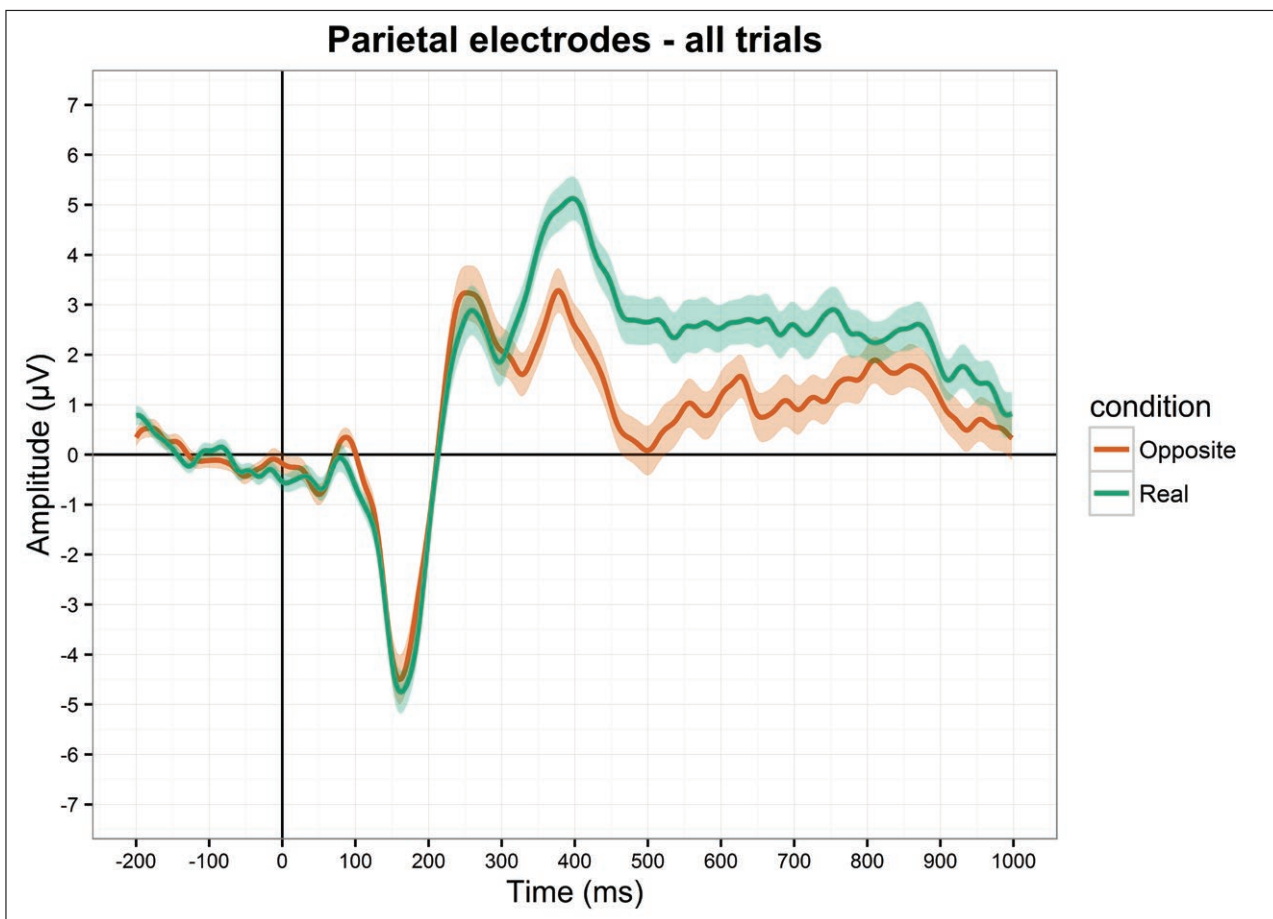


Figure 7: ERPs from all test round trials at the parietal electrodes.

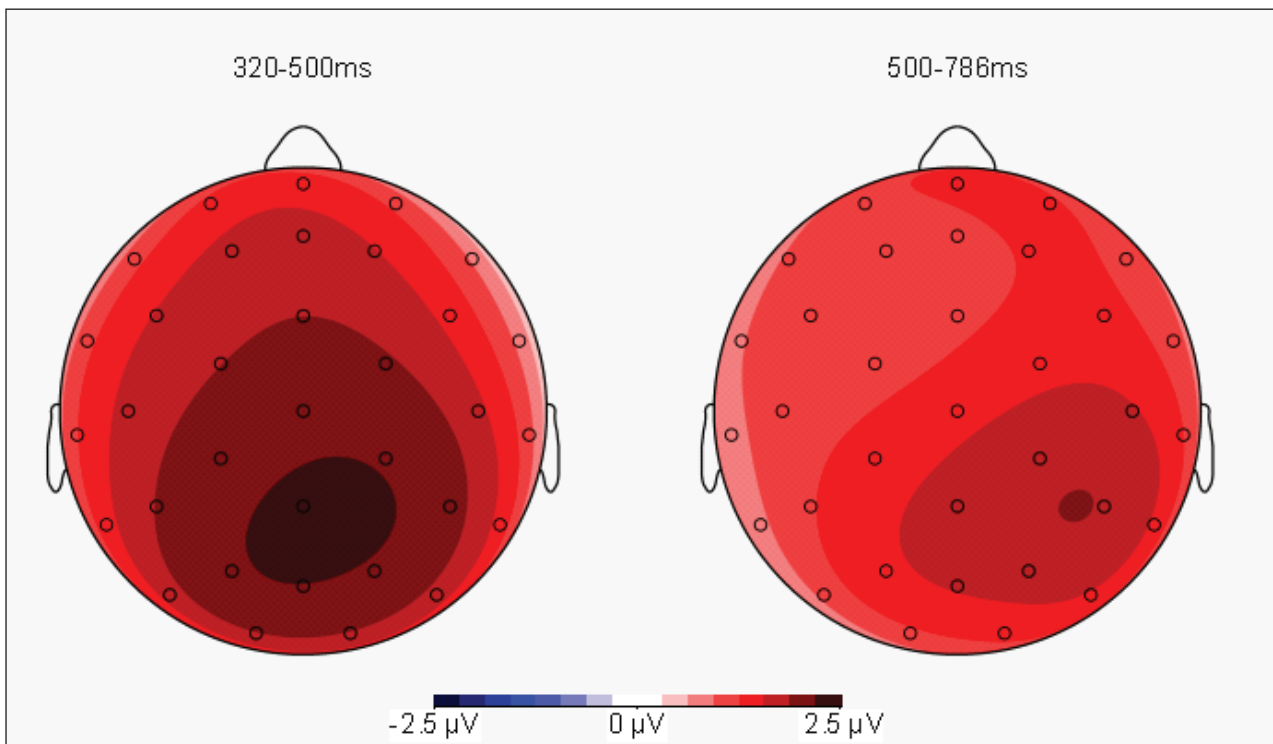


Figure 8: Topographic plots of the real minus opposite difference wave in the test round.

not significant, the P3 effect was greatest in parietal areas in the posterior quadrants.

This analysis included all trials, regardless of whether the participants answered them correctly. To double check, we also analysed only trials which participants answered correctly. Across the 29 participants, an additional 18.6% of trials were rejected due to incorrect responses. Statistical analyses revealed similar results to the analyses of all trials, but all effects were weaker due to having fewer trials.

The effect appears to be centro-parietal according to the topoplots in **Figure 9**, and therefore it is unlikely that lateralisation of language function due to handedness would make any difference to the data. However, we repeated the analyses when excluding the five left-handed participants in the data to double check. Statistical analyses revealed similar results to the analyses of all participants, but all effects were weaker due to having fewer trials.

Accordingly, all statistics reported in the rest of the paper include all trials and all participants.

These analyses are summarised in **Table 2**. Here, *ges* refers to the generalised eta squared measure of effect size.

Correlations between behavioural and neurophysiological results

The ERP difference between conditions during the test round could be driven by the sound-symbolic nature of the ideophones, but it could also be an unrelated learning or memory effect. To tease the two apart, we ran individual differences ranked correlations between ERP results and our two behavioural measures: differences in test scores across conditions in the learning task, and accuracy in the sound-symbolic sensitivity task.

If the P3 amplitude in this experiment is related to how easy the ideophones were to learn in the real versus the opposite condition, then the average P3 amplitude per condition per participant should correlate with the participant's test score in that condition in the learning task. However, there was no correlation between P3 amplitude and test score in the real condition or in the opposite condition, which suggests that the ERP effect may be related to something other than ease of learning or recognition.

The P3 amplitude difference between conditions may instead reflect the participants' sensitivity to sound symbolism. If so, then participants who were more sensitive to sound symbolism—as measured in the separate

sound-symbolic sensitivity check—should show a greater difference between P3 amplitude peaks than participants who were less sensitive to sound symbolism.

We calculated the P3 effect magnitude by subtracting the average amplitude in the opposite condition from the average amplitude in the real condition per participant. We then correlated the effect magnitudes with participants' two-alternative forced choice accuracy scores from the sound-symbolic sensitivity check. These measures were significantly correlated ($r = 0.42, p = 0.0236$), meaning that participants who are better at guessing the meanings of ideophones show a greater P3 effect.

Since the two-alternative forced choice task was significantly correlated with the test score difference between conditions, we also correlated test score differences with P3 amplitude differences across participants. This suggested the same relationship, but was not significant ($r = 0.34, p = 0.067$).

Taken together, the correlations between behavioural measures of sound-symbolic sensitivity and P3 amplitude difference between conditions suggests that the P3 effect found in this experiment is related to an individual's sensitivity to sound symbolism. The lack of a relationship between the P3 amplitude and test score per condition goes some way towards ruling out a non-sound-symbolic learning or recognition effect.

To explore this further, below are plotted the same ERPs for participants grouped according to their score in the two-alternative forced choice task. The top half of participants all scored above the mean of 72.96% ($N = 15$), and the mean of their scores was 79.65%. The bottom half of participants all scored below the mean of 72.96% ($N = 14$), and the mean of their scores was 65.79%. Despite the bottom half of participants still scoring comfortably above chance in the sound-symbolic sensitivity task, the P3 effect from the learning task all but disappeared, as shown in **Figure 9** and **Figure 10**.

When we re-run the ERP ANOVAs for the 2AFC top half and 2AFC bottom half groups separately, the effect is much smaller, indeed not significant, for the 2AFC bottom half group ($F = 4.13, p = 0.063, ges = 0.019$) while remaining consistent for the 2AFC top half group ($F = 14.30, p = 0.0020, ges = 0.12$). There were no factors like age, gender, education, handedness, or number of other languages spoken which may have driven this divide between participants.

| Cluster-based permutation test | All trials | | | Correct answers only | | |
|--------------------------------|---------------------------|---------|-------|--------------------------|--------|-------|
| | 320–786ms $p = 0.0027$ | | | 380–594ms $p = 0.011$ | | |
| ANOVA window | F | p | ges | F | p | ges |
| 320–500ms | 16.99 | 0.00030 | 0.056 | 7.96 | 0.0087 | 0.037 |
| 500–786ms | 8.96 | 0.0057 | 0.032 | 7.86 | 0.0091 | 0.033 |
| 320–500ms (LH removed) | 15.47 | 0.00066 | 0.060 | 5.84 | 0.024 | 0.032 |
| 500–786ms (LH removed) | 5.12 | 0.033 | 0.021 | 3.30 | 0.082 | 0.015 |

Table 2: Table of main effect of condition results.

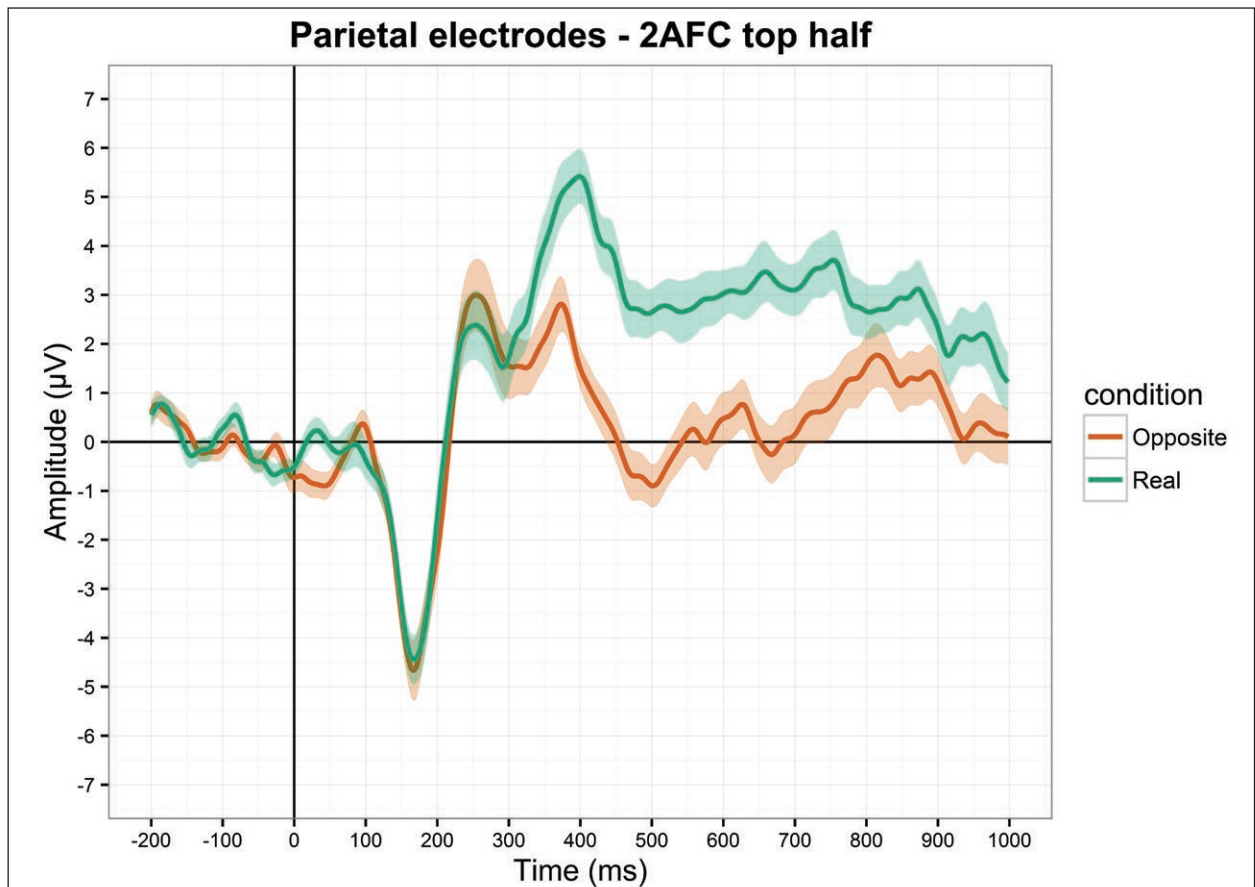


Figure 9: ERPs for the top 15 participants in the 2AFC task measuring sound-symbolic sensitivity.

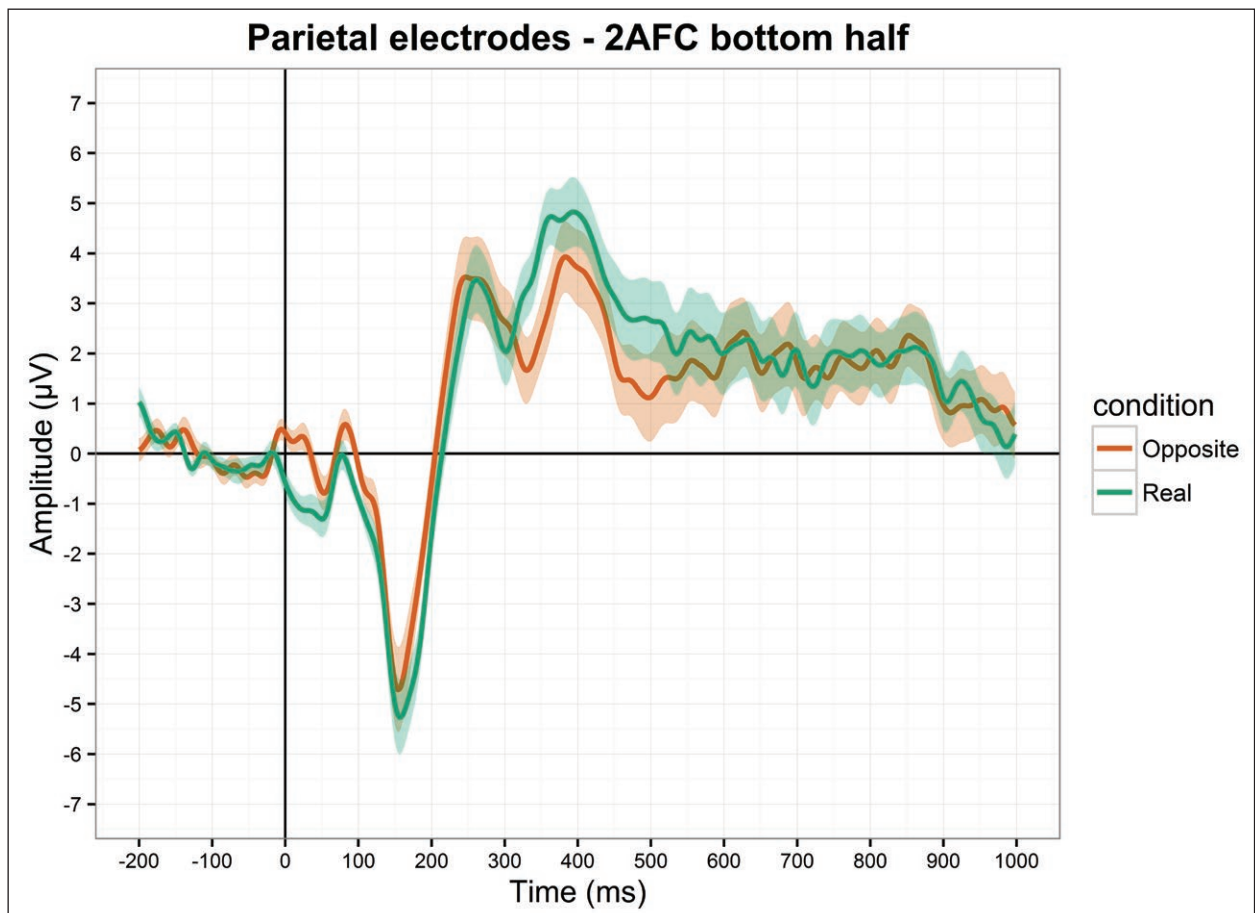


Figure 10: ERPs for the bottom 14 participants in the 2AFC task measuring sound-symbolic sensitivity.

Comparing **Figures 9** and **10** shows that the P3 peak for ideophones in the real condition remains consistent at approximately $5\mu\text{V}$. The difference between the two groups is the amplitude of the P3 peak for ideophones in the opposite condition. For the 2AFC top half group, the opposite P3 peak remains consistent, but for the 2AFC bottom half group, the opposite P3 peak rises to $4\mu\text{V}$. This means that the greater P3 effect in the test round for participants who scored higher in the 2AFC task is driven by the ERPs in response to ideophones learned in the opposite condition, not ideophones learned in the real condition. In the test round, participants in the 2AFC top half group got 88.07% in the real condition and 67.72% in the opposite condition, while participants in the 2AFC low group got 85.15% in the real condition and 75.17% in the opposite condition. This reflects the correlation between sound-symbolic sensitivity and test difference score as shown in **Figure 7**, and provides a useful marker for more extensive behavioural experiments with a larger sample size.

Discussion

Dutch speakers process Japanese ideophones differently, both behaviourally and neurologically, depending on whether they have learned the words with sound-symbolically matching or mismatching meanings, despite not knowing about the manipulation.

Behaviourally, we found that participants learned the sound-symbolically matching word pairs (i.e. the ideophone and its real translation) better than the sound-symbolically mismatching or non-matching word pairs (i.e. the ideophone and its opposite translation). We also found that, despite doing the learning task, participants were still able to guess the meanings of ideophones at above chance accuracy in a two-alternative forced choice test afterwards. Finally, there was a strong correlation between accuracy and reaction times; the more accurately answered ideophones were answered more quickly. All these behavioural findings closely replicate Lockwood et al. [11], and are consistent with the sound-symbolic bootstrapping effect found in learning tasks [27, 28].

In the ERP results, we found no effect of sound symbolism in the learning round, which we speculate is because participants were focused on the learning task; it is possible that effects would arise in a simple judgement or priming task. We did find an effect in the test round, where the presence (or absence) of sound symbolism influenced the amplitude of the P3 and late positivity. The P3 amplitudes per condition did not correlate with participants' test scores per condition, which suggests that the effect is not simply due to ease of learning. However, we did find that the P3 effect magnitude correlated with the behavioural measures of sound-symbolic sensitivity in the 2AFC task performed after the main experiment, which suggests that the P3 effect is related to an individual's sensitivity to sound symbolism.

To further explore this effect, we looked at individual differences between participants and found a relationship between the ERP results and the two behavioural measures of sound symbolism: performance in the

sound-symbolic sensitivity check and differences between test scores across conditions. We found that the magnitude of the ERP effects correlated with the performance in the behavioural tasks and thus serves as an index of sound-symbolic sensitivity. This was not hypothesised *a priori*, but the finding provides additional evidence that sound-symbolic sensitivity affects word learning and recognition. It is worth stressing the fact that the behavioural measures from a task measuring sound-symbolic sensitivity predict the ERPs from a completely separate learning and test task (which was done before the participants did the 2AFC task); this suggests that sound-symbolic sensitivity is a consistent process or state which affects how well participants learn sound-symbolic words. To our knowledge, this is the first report of individual differences in sound-symbolic learning and decision tasks being correlated to neurophysiological measures. Rather than noise to be averaged out, these differences can be used to zoom in on the causal processes underlying sound-symbolism and iconicity.

The ERP findings partially mirror existing work on Japanese ideophones and ERPs by Lockwood and Tuomainen [29], who found that ideophones elicited a greater late positive complex than arbitrary words. Two factors make it difficult to confidently draw functional interpretations for each component from this data; the fact that the P3 and late positive complex are related to all kinds of different functional roles, and the fact that neuroimaging research on sound symbolism in real words is in its infancy. However, we provide two possible interpretations here.

Firstly, the P3 is greater in response to the ideophones learned in the real condition. The P3 is a well-documented component related to attention, and has been functionally separated into a frontal P3a broadly related to stimulus novelty and a parietal P3b broadly related to memory processes [30]. The latency and topographic distribution of the effect here suggests that it is a P3b, whose amplitude varies with task demands; increases in memory load reduce P3 amplitude because of the greater task processing demands. Individual difference measures suggest that the P3 effect was related to sound symbolism per condition rather than ease of learning and recognition per condition. Moreover, it appears that the reason for the increased difference in P3b amplitudes between conditions is due to variation in P3b amplitude to ideophones learned in the opposite condition. Participants who scored above the mean in the 2AFC task had a lower P3b amplitude in response to ideophones in the opposite condition; participants who scored below the mean in the 2AFC task had a higher P3b amplitude to ideophones learned in the opposite condition. Coupled with the fact that participants who scored higher in the 2AFC task had a greater accuracy difference between the test scores for ideophones in each condition, this suggests that all participants found learning sound-symbolic words similarly easy, but being more sensitive to sound symbolism makes it harder to learn non-sound-symbolic words and requires extra resource allocation. Therefore, we theorise that the P3 amplitude can be used as an index of the degree to

which an individual participant must suppress conflicting cross-modal information during learning and recognition.

In future studies, we would expect to see an even greater P3 amplitude difference between conditions in a similar experiment with pseudowords which deliberately maximise attested cross-modal contrasts. Since eliciting the P3 requires a response during a match/mismatch paradigm [31], it is perhaps unsurprising that we do not find a P3 effect in the initial passive learning rounds.

Secondly, the late positive complex is also greater in response to the ideophones learned in the real condition. The late positive complex in language tasks is generally linked to increased complexity [32], working memory demands [33], or violation of expectation [6], although there is a lot of individual variation [34]. It has also been linked to emotionally arousing stimuli (and referred to as the late positive potential) in non-language ERP literature [35, 36, 37]. It is possible that the late positive potential observed here and in Lockwood and Tuomainen [16] are more like those found in the emotion literature. Ideophones are frequently described as being vivid or synaesthetic in how they express meaning, and are particularly well-suited to conveying affective states [20, 38]. Perhaps the late positive potential elicited by ideophones in Lockwood and Tuomainen [16] and by ideophones with their real meanings in this study is an indication of their emotional or attentional salience in comparison to arbitrary words or words without sound-symbolic associations. However, there may be a simpler explanation: the strong correlation between P3 effect magnitude and late positive complex effect magnitude ($r = 0.46$, $p = 0.0124$) suggests that the two components overlap to the extent that the observed late positive complex in this experiment is just a continuation of the large P3 effect, not a separate component reflecting a separate process.

One limitation of the current study is that the stimuli were not counterbalanced across participants. However, we found in pre-tests with the same counterbalanced stimuli in Lockwood et al. [11] that the behavioural learning effect was consistent for both groups. Another caveat is that the individual difference data is exploratory and should not be taken as conclusive.

Conclusion

Dutch speakers are sensitive to the meanings of Japanese ideophones. Ideophones with their real translations are learned more effectively than ideophones with their opposite meanings due to the congruent cross-modal associations which sound symbolism provides. These associations are accessible despite the learning task, as ideophones were still accurately guessed in a two-alternative forced choice task which took place after the learning task.

Moreover, performance in the 2AFC task actually predicted learning differences between conditions and P3 effect magnitude. This confirms that sound symbolism boosts word learning in adults learning words in a new language, in addition to existing evidence from infants and children as well as adults. It also provides evidence that sound-symbolic cues in Japanese ideophones are available to speakers of an unrelated language, suggesting

a fruitful avenue of research into the universality of sound-symbolic cues in ideophones across languages [14]. While the word learning task is not fully representative of language in a natural context – it is almost impossible to marry full experimental control with full ecological validity – it does go further than the forced choice experiments with pseudowords which make up the majority of sound symbolism research.

Our results pave the way for future work further unravelling the neural correlates and time course of sound symbolism, and suggest that the P3 is heavily implicated in sound symbolism. We suggest that the P3 amplitude is an index of the degree to which the sounds of a word cross-modally match the word's sensory meaning, and that individual differences in sound-symbolic sensitivity constitute a promising inroad for charting the cognitive processes involved in sound symbolism.

Supplementary Materials

All the stimuli, presentation materials, participant data, and analysis scripts can be found on this paper's project page on the Open Science Framework.

<https://osf.io/ema3t/>

Acknowledgements

This work was funded by an International Max Planck Research Scholarship awarded to Gwilym Lockwood and by a Veni grant from NWO (Netherlands Organisation for Scientific Research) awarded to Mark Dingemanse. We thank Atsuko Takashima for providing the recordings, Arne Driessen for helping with the translations, and Ashley Lewis for his help with the cluster-based permutation tests.

Competing Interests

The authors declare that they have no competing interests.

References

1. Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. 2015. Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* 19(10): 603–615. DOI: <http://dx.doi.org/10.1016/j.tics.2015.07.013>
2. Perniss, P., and Vigliocco, G. 2014. The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651): 20130300. DOI: <http://dx.doi.org/10.1098/rstb.2013.0300>
3. Lockwood, G., and Dingemanse, M. 2015. Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Language Sciences* 1246. DOI: <http://dx.doi.org/10.3389/fpsyg.2015.01246>
4. Perniss, P., Thompson, R. L., and Vigliocco, G. 2010. Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* 1(E227). DOI: <http://dx.doi.org/10.3389/fpsyg.2010.00227>
5. Dingemanse, M. 2012. Advances in the Cross-Linguistic Study of Ideophones. *Language and*

- Linguistics Compass 6(10): 654–672. DOI: <http://dx.doi.org/10.1002/Inc3.361>
6. Voeltz, F. K. E., and Kilian-Hatz, C. 2001. Ideophones. Amsterdam: John Benjamins Publishing.
 7. Oda, H. 2000. An embodied semantic mechanism for mimetic words in Japanese. Ph.D. dissertation, Indiana University, Indiana.
 8. Iwasaki, N., Vinson, D. P., and Vigliocco, G. 2007. What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese-Language Education around the Globe* 17: 53–78.
 9. Kunihiro, S. 1971. Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior* 10(4): 427–429. DOI: [http://dx.doi.org/10.1016/S0022-5371\(71\)80042-7](http://dx.doi.org/10.1016/S0022-5371(71)80042-7)
 10. Nygaard, L. C., Cook, A. E., and Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition* 112(1): 181–186. DOI: <http://dx.doi.org/10.1016/j.cognition.2009.04.001>
 11. Lockwood, G., Dingemans, M., and Hagoort, P. 2016. Sound-Symbolism Boosts Novel Word Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. DOI: <http://dx.doi.org/10.1037/xlm0000235>
 12. Akita, K. 2009. A Grammar of Sound-Symbolic Words in Japanese: Theoretical Approaches to Iconic and Lexical Properties of Mimetics. Kobe University.
 13. Hamano, S. 1998. The Sound-Symbolic System of Japanese. Center for the Study of Language and Information ; Tokyo.
 14. Dingemans, M., Schuerman, W., Reinisch, E., Tufvesson, S., and Mitterer, H. 2016. What Sound Symbolism Can and Cannot Do: Testing the Iconicity of Ideophones from Five Languages. *Language*, 92(2): e67–83.
 15. Samarin, W. 1965. Perspective on African ideophones. *African Studies* 24(2): 117–21.
 16. Lockwood, G., and Tuomainen, J. 2015. Ideophones in Japanese modulate the P2 and late positive complex responses. *Frontiers in Psychology*, 6, 933. DOI: <http://dx.doi.org/10.3389/fpsyg.2015.00933>
 17. Revill, K. P., Namy, L. L., DeFife, L. C., and Nygaard, L. C. 2014. Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language* 128(1): 18–24. DOI: <http://dx.doi.org/10.1016/j.bandl.2013.11.002>
 18. Kanero, J., Imai, M., Okuda, J., Okada, H., and Matsuda, T. 2014. How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words. *PLoS ONE* 9(5): e97905. DOI: <http://dx.doi.org/10.1371/journal.pone.0097905>
 19. Duffloth, G. 1972. Notes on expressive meaning. *Chicago Linguistic Society* 8: 440–447.
 20. Kita, S. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35(2), 379–415. DOI: <http://dx.doi.org/10.1515/ling.1997.35.2.379>
 21. Kunene, D. 1965. The Ideophone in Southern Sotho. *Journal of African Languages* 4: 19–39.
 22. Kovic, V., Plunkett, K., and Westermann, G. 2010. The shape of words in the brain. *Cognition* 114(1): 19–28. DOI: <http://dx.doi.org/10.1016/j.cognition.2009.08.016>
 23. Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., and 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. DOI: <http://dx.doi.org/10.1016/j.bandl.2015.03.011>
 24. Levinson, S. C. 2012. The Original Sin of Cognitive Science. *Topics in Cognitive Science* 4(3): 396–403. DOI: <http://dx.doi.org/10.1111/j.1756-8765.2012.01195.x>
 25. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. 2011. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* 2011: e156869. DOI: <http://dx.doi.org/10.1155/2011/156869>
 26. Weissgerber, T. L., Milic, N. M., Winham, S. J., and Garovic, V. D. 2015. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biol* 13(4): e1002128. DOI: <http://dx.doi.org/10.1371/journal.pbio.1002128>
 27. Imai, M., and Kita, S. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651): 20130298. DOI: <http://dx.doi.org/10.1098/rstb.2013.0298>
 28. Imai, M., Kita, S., Nagumo, M., and Okada, H. 2008. Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65. DOI: <http://dx.doi.org/10.1016/j.cognition.2008.07.015>
 29. Kita, S. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35(2), 379–415. DOI: <http://dx.doi.org/10.1515/ling.1997.35.2.379>
 30. Polich, J. 2007. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* 118(10): 2128–2148. DOI: <http://dx.doi.org/10.1016/j.clinph.2007.04.019>
 31. Key, A. P. F., Dove, G. O., and Maguire, M. J. 2005. Linking Brainwaves to the Brain: An ERP Primer. *Developmental Neuropsychology* 27(2): 183–215. DOI: http://dx.doi.org/10.1207/s15326942dn2702_1
 32. Lau, E. F., Phillips, C., and Poeppel, D. 2008. A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience* 9(12): 920–933. DOI: <http://dx.doi.org/10.1038/nrn2532>
 33. Friederici, A. D., Steinhauer, K., Mecklinger, A., and Meyer, M. 1998. Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology* 47(3): 193–221.
 34. Kos, M., van den Brink, D., and Hagoort, P. 2012. Individual Variation in the Late Positive Complex to

- Semantic Anomalies. *Frontiers in Psychology* 3. DOI: <http://dx.doi.org/10.3389/fpsyg.2012.00318>
35. Hajcak, G., Dunning, J. P., and Foti, D. 2009. Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology* 120(3): 505–510. DOI: <http://dx.doi.org/10.1016/j.clinph.2008.11.028>
36. Hajcak, G., MacNamara, A., and Olvet, D. M. 2010. Event-Related Potentials, Emotion, and Emotion Regulation: An Integrative Review. *Developmental Neuropsychology* 35(2): 129–155. DOI: <http://dx.doi.org/10.1080/87565640903526504>
37. Moratti, S., Saugar, C., and Strange, B. A. 2011. Prefrontal-Occipitoparietal Coupling Underlies Late Latency Human Neuronal Responses to Emotion. *The Journal of Neuroscience* 31(47): 17278–17286. DOI: <http://dx.doi.org/10.1523/JNEUROSCI.2917-11.2011>
38. Doke, C. M. 1948. The Basis of Bantu Literature. *Africa*, 18(4), 284–301. DOI: <http://dx.doi.org/10.2307/3180450>

Peer review comments

The author(s) of this paper chose the Open Review option, and the peer review comments are available at: <http://dx.doi.org/10.1525/collabra.42.opr>

How to cite this article: Lockwood, G, Hagoort, P and Dingemans, M 2016 How Iconicity Helps People Learn New Words: Neural Correlates and Individual Differences in Sound-Symbolic Bootstrapping. *Collabra*, 2(1): 7, pp. 1–15, DOI: <http://dx.doi.org/10.1525/collabra.42>

Submitted: 06 April 2016 **Accepted:** 25 May 2016 **Published:** 06 July 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.