To appear in JASA

The alveolar trill is perceived as jagged/rough by speakers of different languages

Aleksandra Ćwiek,[1] Rémi Anselme,[2] Dan Dediu,[3,4,5] Susanne Fuchs,[1,6] Shigeto Kawahara,[7] Grace E. Oh,[8] Jing Paul,[9] Marcus Perlman,[10,a] Caterina Petrone,[11] Sabine Reiter,[12] Rachid Ridouane,[13] Jochen Zeller,[14] and Bodo Winter[10]

[1] *Leibniz-Centre General Linguistics, Berlin, 10719, Germany*

[2] *Laboratoire Dynamique Du Langage UMR 5596, Université Lumière Lyon 2, Lyon, 69363, France*

[3] *Department of Catalan Philology and General Linguistics, University of Barcelona, Barcelona, 08007, Spain*

[4] *Universitat de Barcelona Institute of Complex Systems (UBICS), Barcelona, 08038, Spain*

[5] *Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, 08010, Spain*

[6] *IMéRA Institute for Advanced Studies, Marseille, 13004, France*

[7] *The Institute of Cultural and Linguistic Studies, Keio University, Mita Minatoku, Tokyo, 108-8345, Japan*

[8] *Department of English Language and Literature, Konkuk University, Seoul 05029, South Korea*

[9] *Asian Studies Program, Agnes Scott College, Decatur, GA 30030, USA*

[10] *Department of Linguistics and Communication, University of Birmingham, Birmingham, B15 2TT, United Kingdom*

[11] *Aix-Marseille Université, CNRS, Laboratoire Parole et Langage, 13100 Aix-en-Provence, France*

[12] *Depto. de Polonês, Alemão e Letras Clássicas, Universidade Federal do Paraná, 80060-150 Curitiba, Brazil*

[13] *Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS and Sorbonne Nouvelle, 75005 Paris, France*

[14] *School of Arts, Linguistics Discipline, University of KwaZulu-Natal, Durban 4041, South Africa*

This paper is part of a special issue on Iconicity and Sound Symbolism.

---

[a] Email: m.perlman@bham.ac.uk

Abstract

Typological research shows that across languages, trilled [r] sounds are more common in adjectives describing rough as opposed to smooth surfaces. We build on this lexical research with an experiment with speakers of 28 different languages from 12 different families. Participants were presented with images of a jagged and a straight line and imagined running their finger along each. They were then played an alveolar trill [r] and an alveolar approximant [l] and matched each sound to one of the lines. Participants showed a strong tendency to match [r] with the jagged line and [l] with the straight line, even more consistently than in a comparable cross-cultural investigation of the *bouba/kiki* effect. The pattern is strongest for matching [r] to the jagged line, but also very strong for matching [l] to the straight line. While we found this effect with speakers of languages with different phonetic realizations of the rhotic sound, it was weaker when trilled [r] was the primary variant. We suggest that when a sound is used phonologically to make systemic meaning contrasts, its iconic potential may become more limited. Our findings extend our understanding of iconic cross-modal correspondences, highlighting deep-rooted connections between auditory perception and touch/vision.

I. **INTRODUCTION**

There is a large amount of research on sound symbolism, documenting how people often attribute meaning to speech sounds (Lockwood and Dingemanse, 2015). For example, experiments with speakers from different languages show that the high front vowel [i] is associated with the meaning of smallness, compared to low back vowels (Auracher, 2017; Hoshi et al., 2019; Knoeferle et al., 2017; Newman, 1933; Parise and Spence, 2012; Sapir, 1929; Tarte and Barritt, 1971). This pattern is hypothesized to stem from the fact that the high second formant frequency and large dispersion of the first and second formant frequencies of [i] correspond to the acoustics of small resonators (Fitch, 1994; Ohala, 1983; Winter et al., 2021). Importantly, this pattern has not only been attested in experiments, but is reflected in vocabulary structure across languages, with high-front vowels occurring more frequently in words denoting smallness (Blasi et al., 2016; Fitch, 1994; Haynie et al., 2014; Huang et al., 1969; Johansson et al., 2019; Johnson, 1967; Levickij, 2013; Thorndike, 1945; Ultan, 1978; Winter and Perlman, 2021). This evidence from lexical and experimental studies is understood as a case of iconicity – a resemblance between the form of a signal (e.g., a word, gesture, or sign) and its meaning. A growing number of scholars argue that iconicity is a fundamental property of languages, spoken and signed (Dingemanse et al., 2015; Perniss et al., 2010).

In cases such as the association between vowels and size, iconicity is *crossmodal*, mapping between sound and qualities that are primarily related to different sensory modalities. Perhaps the most famous example of crossmodal iconicity is the *bouba*/*kiki* effect, where nonce words like *bouba* (or *maluma*) are matched to round shapes, as opposed to nonce words like *kiki* (or *takete*), which are matched to angular shapes (Köhler, 1929; Ramachandran and Hubbard, 2001). This association has been experimentally demonstrated across cultures with speakers of a large set of genealogically diverse spoken languages (Bremner et al., 2013; Ćwiek et al., 2022), and observational studies have found that roundness/angularity is statistically associated with *bouba*- and *kiki*-like speech sounds in the lexicon of English (Sidhu et al., 2021). Experimental evidence suggests that multiple analogies may underpin the perceived resemblance between *bouba*/*kiki* and round/angular shapes, including mediation through emotional arousal (Aryani et al., 2020), and through the similarity between the word *bouba* and the sounds produced by falling or bouncing round objects as opposed to angular ones (Fort and Schwartz, 2022).

The current study focuses on another case of crossmodal iconicity that may exert an influence on the phonological shape of words: the association of rhotic consonants with rough texture. An early study asking American English speakers to rate the qualities of speech sounds found that /r/ was judged as rougher than other phonemes (Greenberg and Jenkins, 1966). In line with this result, a cross-linguistic analysis of poetic texts found that /r/ was over-represented in poems with aggressive rather than tender tone (Fónagy, 1961). It is worth noting that in those studies, it remains unclear which specific realizations of the phoneme (as [r], [ɹ], [ʀ], or another speech sound) the conclusions are drawn from. Recently the association between /r/ and roughness has been found to be widespread across spoken vocabularies. Winter et al. (2022) first showed that for a set of 100 English adjectives rated for roughness (e.g., *rough, abrasive, prickly, smooth, coarse, cottony, silky, oily*), the rhotic phoneme is statistically associated with descriptors of rough surfaces. This pattern was also found across 38 other Indo-European languages and replicated in Hungarian, a Uralic language. In a typological analysis of vocabulary data from lexical databases, trilled /r/ sounds, as indicated in the phonologically coded lexical databases, were found to be much more common in translational equivalents of 'rough' rather than 'smooth' for a diverse sample of 332 spoken languages from 84 phyla (see also Levickij, 2013). Considering that perceptual studies of surface touch suggest that the spatial frequency of grating patterns is a primary determinant of textural roughness (Hollins and Bensmaïa, 2007; Lederman, 1974, 1983), and that spatial frequency is perceptually associated with auditory amplitude modulations (Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013; Sherman et al., 2013), Winter et al. (2022) suggested that the intermittent tongue movements of trills and the resulting repetitive amplitude modulations (see Fig. 1a) might provide the iconic motivation behind this pattern.

However, a recent study by Anselme et al. (2022) calls into question whether the statistical association with rough meanings in vocabulary data is specific to the trilled /r/ phoneme, or whether it might be associated with rhotic cononants more broadly. Anselme et al. (2023) provide evidence that the precise phonetic realization of rhotics has not always been accurately documented in the data bases used by Winter et al. (2022): although /r/ technically symbolizes an alveolar trill according to the International Phonetic Alphabet, it is often used to represent a generic "r-like" sound, not reliably distinguishing whether it is typically realized as a trill. Anselme et al. (2022) recoded a substantial portion of Winter et al.'s (2022) typological data, and in their re-analysis found that "r-like" sounds in general, not just trills, are associated with roughness across spoken language lexicons. Thus, it is not entirely clear whether the patterns found in Winter et al.'s analysis of lexical data is rooted in an iconic association between roughness and /r/ realized specifically as an alveolar trill, or whether it is driven by rhotics more generally, regardless of how they are phonetically realized.

Taken together, evidence from lexical databases suggests that the trilled phoneme /r/ is associated with roughness (Winter et al., 2022). However, the overuse of the IPA symbol /r/ to represent in writing various r-like sounds without, in many cases, specifying unambiguously which particular speech sounds it stands for, prevents us from concluding that it is specifically the trill that bears this semantic association (Anselme et al., 2022). The current study addresses this ambiguity by directly testing the connection between specifically the alveolar trill [r] realized in a controlled and explicit manner, and roughness across a diverse language sample in order to explore the cross-linguistic potential of the hypothesized association. We follow up on the lexical pattern found by Winter et al. (2022) with a perception experiment to assess whether the alveolar trill [r] is perceived as rough by speakers of 28 different languages. We tested [r] against the lateral [l], another liquid with an alveolar place of articulation, but with no strong repetitive amplitude modulations (see Fig. 1b). By presenting acoustic stimuli to speakers of a diverse set of languages, we were also able to assess the extent to which speakers of languages with different phonetic realizations of rhotic consonants differ in their crossmodal associations of the alveolar trill [r]. Given that phonemes primarily serve a contrastive function to distinguish words within languages, it is possible that speakers of a language that uses an alveolar trill [r] as the primary phonetic variant may treat this phoneme as relatively more "arbitrary," and less imbued with meaning. Therefore, we wanted to assess whether having the alveolar trill [r] as the primary allophone of the /r/ phoneme in one's language could potentially diminish the strength of its iconic association. Similarly, we were able to explore whether distinguishing /r/ and /l/ sounds phonemically in one's grammar also plays a role in modulating the perceived crossmodal iconicity of the alveolar trill [r].

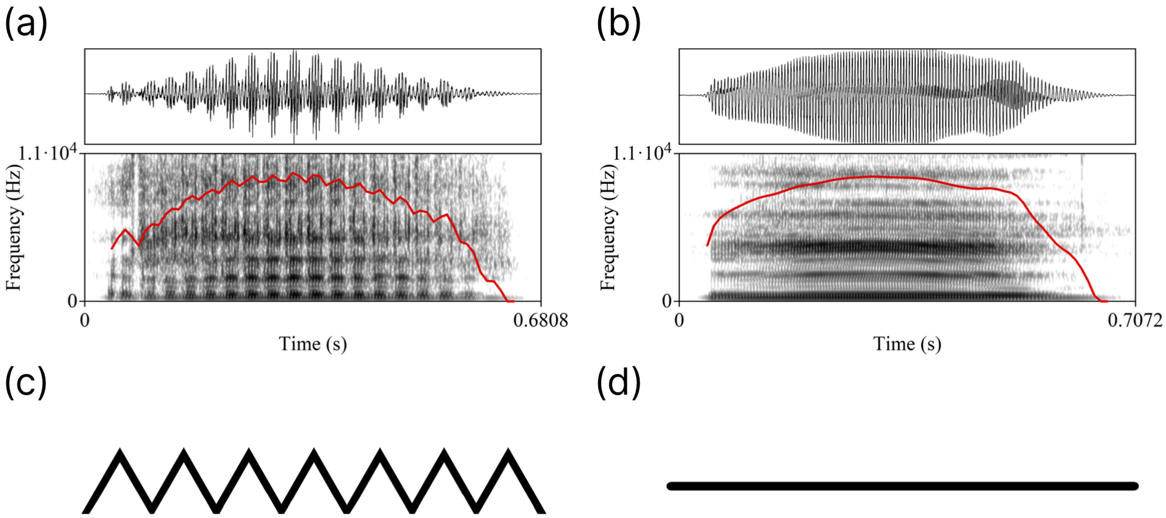

FIG 1. The oscillograms and spectrograms for the recording of (a) the alveolar trill [r], and (b) the alveolar lateral approximant [l]. The superimposed red line is the intensity curve with a range between 55 and 85 dB. The jagged line (c) and the flat line (d) were the corresponding visual stimuli presented to participants in the experiment.

In our experiment, we used the shapes shown in Fig. 1c and 1d as visual stimuli, asking participants to imagine what it feels like to touch these surfaces. Thus, the connection specifically between touch and sound, which we were seeking to investigate, is indirect with these visually presented stimuli, despite highlighting haptic touch to our participants via the instructions. However, the use of visual stimuli rather than felt surfaces was necessary to conduct the experiment online (see below), which prevented the use of textural stimuli. In selecting these visual shapes as representations of textures, the contrast between a jagged and a flat shape was motivated by studies suggesting that the frequency of spatial grating predicts roughness (Hollins and Bensmaïa, 2007; Lederman, 1974, 1983). Furthermore, surface texture is what is known as a "common sensible," a percept that can be perceived through multiple different modalities (Marks, 1978). Roughness in particular can also be perceived via vision (Lederman and Abbott, 1981) and audition (Lederman, 1979), and has similar psychometric functions in these modalities. Nevertheless, our experiment is somewhat ambiguous with respect to vision and touch, testing a stimulus that can be perceived either as "jagged," relating to the construct of "shape," or as "rough," relating to the construct of "texture."

II. METHODS

The experiment reported here was conducted as a part of a larger study (Ćwiek et al., 2021, 2022), which included both an online web experiment and an on-site field experiment. Our overarching goal for conducting the same experiment online and in the field was to maximize linguistic and cultural diversity, with the goal of targeting non-WEIRD (Western Educated Industrialized Rich Democratic) communities (Blasi et al., 2022; Henrich et al., 2010). Participation in the web experiment required literacy, as well as access to and experience with the internet. The experiment conducted on site did not require participants to be literate, and thus it allowed us to target speakers with limited formal education as well as limited access to the internet and globalized culture.

As the web and field experiment differ only slightly (see Procedure) and are also analyzed in the same statistical model for ease of presentation (see Statistical Analysis), we treat them as two separate samples from the same study. All of the data and code for the experiments are available in an Open Science Framework repository, at: https://osf.io/mjcnq/

A. Participants

We used opportunity sampling for both the web experiment and the field experiment. All participants indicated their informed consent and completed the study on a volutary basis.

For the web experiment, we distributed the survey online via social media or via directly contacting native speakers and asking them to share the link to the experiment with their friends and family. Our initial convenience sample for the web experiment included data from 975 participants. We excluded participants who indicated they did not speak the language of the survey (*n*=9), who failed to provide both responses (*n*=38), or who selected a response without playing back the sound (*n*=22). Additionally, we did not obtain enough Tamily and Malagasy data (one and two speakers respectively) for them to be included in our analysis. In total, we excluded 72 participants (7.4%) from the analysis, leading to final sample with data from 903 participants representing 25 languages from 9 language families, as detailed in Table I. 781 participants (86.5%) spoke a second language and 122 participants (13.5%) self-reported to be monolingual. Of the participants who were not native English speakers, 727 (84.1%) spoke English as a second language. In terms of gender composition, our sample included 681 female speakers (75.4%) and 222 male speakers (28.3%). Participants ranged from 18 to 84 years of age (mean 32.9, median 29).

For the field experiment, opportunity sampling involved collaborating with linguists that were going on field visits during the period of the study. The field experiment was conducted on 6 sites, with a total of 133 participants who were speakers of 6 different languages from 4 families, including Palikúr, Brazilian Portuguese, Daakie, Tashlhiyt, German, and English (see Table II). Four of these language groups (Palikúr, Brazillian Portuguese, Daakie) were targeted as non-WEIRD communities, with limited formal education and access to the internet and globalized culture. Palikúr data were collected at the banks of Oyapock river near St. Georges de l'Oyapock in French Guayana (at the border with Brazil). Brazilian Portuguese data were collected with a quilombo community from the Cametá region in Brazil. Both Palikúr and Brazilian Portuguese speakers live in Amazonia and are rural communities of farmers/hunters who sell their goods on the market. Daakie data were collected with a farming/hunting community living in Port Vato on Ambrym, Vanuatu. All three communities do not have regular access to electricity, and the use of mobile phones is highly limited because of lacking resources and connection service. Access to education is limited in these communities. For comparison, English, German, and Berber speakers were recruited so that we could disentangle the effects of task (web experiment versus field experiment) from characteristics of the participant sample. English and Tashlhiyt data were collected among university students in Birmingham, UK, and Agadir, Morocco, respectively. German data were collected among residents of a holiday resort in Lubmin, Germany. The specific setting and participant sample for the field experiment differed across the six language groups, reflecting the various on-site conditions. The Daakie speakers took part in the study in a small concrete building belonging to the Presbyterian Church, seated at a table on a bench, with efforts made to minimize distractions from bystanders. Brazilian Portuguese participants performed the task in their homes; Palikúr speakers in a communal building where they were interviewed one-on-one in a separate room. English and Tashlhiyt participants performed the task in a quiet room in a university; German speakers in a quiet bungalow in a holiday resort.

Sample size in the field experiment varied based on the availability and willingness of participants on each site. We excluded six participants (4.5%) who failed to provide responses for each sound stimulus, leaving us with a sample of 127 participants. Of these, 75 speakers (59.1%) spoke a second language, and 52 speakers (40.9%) self-reported to be monolingual. Specifically for the three target languages – Palikúr, Brazillian Portuguese, Daakie – the figure of second language speakers was 21 (63.6%), in contrast to 12 monolingual speakers (36.4%). Only 1 participant from the target languages (3.0%) self-reported to know English, as opposed to 32 who did not (97.0%). The final sample included 91 female participants (71.7%) and 36 male participants (28.3%). Ages ranged from 18 to 75 years (mean 28.6, median 20.0).

B. **Materials**

The acoustic stimuli included a recording of the alveolar trill [r] and a recording of the lateral alveolar approximant [l] (see Fig. 1a and 1b). These sound were produced by a native Polish speaker with training in phonetics (author AĆ). The sounds were produced in isolation, without any carrier phrase or vocalic context. The rough and smooth textures were represented with two line drawings, one jagged/rough, one flat/smooth (Fig. 1c and 1d). Participants were instructed to imagine moving their finger along the lines to emphasize the touch dimension.

C. **Procedure**

In addition to the current experiment, the complete study included a main task involving guessing the meaning of novel iconic vocalizations (Ćwiek et al., 2021) and an additional task involving *bouba*/*kiki* (Ćwiek et al., 2022), with the current study always coming last. Thus, these other tasks were both related to different kinds of vocal iconicity and sound symbolism.

Importantly, however, participants were not provided with any feedback on their guessing in either of the previous experiments. For the entire set of studies, we collaborated with native speakers who translated the consent forms and instructions (Ćwiek et al., 2021, 2022). The web experiment was hosted on the Percy platform (Draxler, 2011), and accessed by participants via their personal computer, smartphone, or tablet over the internet. The field experiment was conducted orally in the native language of the participants, including the consent process and all instructions. The consent and the instructions were read to the participants, and they also had opportunity to read these themselves. All participants provided signed consent. For English, German, and Tashlhiyt speakers, this procedure and the experiment were conducted by linguists who were also native speakers. In the case of Daakie and Brazilian Portuguese speakers, this was done by linguists who knew the respective languages. For Palikúr speakers, the field linguist conducted the experiment in Brazilian Portuguese with an on-site interpreter translating into Palikúr.

The task was identical across all languages, but differed slightly between the web and field experiment. In the web experiment, participants were presented with images of the two lines next to each other on a screen. They then listened to each auditory stimulus separately, making a response after hearing each sound (sequential rather than paired matching). The order of presentation of the auditory stimuli, as well as the images (left vs. right), was randomized. For the field experiment, participants were simultaneously presented with both lines and were played both sounds via laptop speakers of the respective experimenter before making their response, enabling paired matching after listening to both sounds. The lines were printed out on white paper in A5 format and presented on a table in front of the participant. In contrast to the web experiment, the presentation order of the line drawings (left vs. right) and the sounds was not recorded and not controlled for.

In the web experiment, participants could click to replay each sound, and in the field experiment, they could ask the experiment to play a sound again. After completing the full study, participants were asked for background information on their sex, age, native language(s), and other known languages – via written questions in the web experiment, and via oral questions in the field experiment. The web experiment additionally asked for the participants' country of residence, and the place where they entered primary school. Additionally, we inquired about the environment in which they completed the survey, the input device and audio output device they used, and their hearing ability.

D. Phonetic coding of rhotics for both samples

To investigate the effect of language background on participants' judgments, we coded what rhotic variant characterized each of the languages spoken by our participants. The coding procedure was based on Anselme et al. (2023) and used resources from large databases with phonemic and phonetic information on the languages spoken by our participants, especially PHOIBLE (Moran et al., 2014) and Glottolog (Hammarström et al., 2020). However, language-specific sources were consulted for each language separately, including recordings such as those available in the DoReCo corpus (Seifart et al., 2022). All the information on the procedure, the sources, and the individual sources consulted can be found in the OSF repository: https://osf.io/mjcnq/

We coded rhotic variants separately for each speaker's first and second languages. There were three dimensions of coding, each binary-coded as occurring (1) or not (0). First, we coded the languages for whether they have a phonemic contrast between /r/ and /l/. Second, we coded whether each language uses an alveolar trill [r] as the main r-sound. Third, we coded whether [r] can feature as an allophone in each language.

When coding foreign languages reported by the participants, we marked a variable as present for this participant if any of the languages they spoke had the variable we were looking for. For example, if a participant reported speaking Polish and German as foreign languages, they would be marked as "1" for "[r] as the main r-sound," as they spoke at least one language in which this was the case. The results of the coding for each language can be found in Tables I and II, for the web and the field experiment, respectively.

TABLE I. Counts of participants per language and language family in the web experiment. The table is ordered alphabetically by language, within family and genus. The participant sample is discussed in Section A; the rhotic coding is discussed in Section D; the “match” variable is discussed in Section E.

Family	Genus	Language	N of participants	r/l contrast	[r] as main r-sound	[r] as allophone	“match”
Atlantic-Congo	Bantu	Zulu	20	0	0	1	85.0%
Indo-European	Albanian	Albanian	10	1	1	1	70.0%
	Armenian	Armenian	20	1	1	1	85.0%
	Germanic	Danish	18	1	0	1	94.4%
		English	39	1	0	1	97.4%
		German	85	1	0	1	95.3%
		Swedish	21	1	0	1	95.2%
	Greek	Greek	42	1	0	1	90.5%
	Iranian	Farsi	21	1	1	1	85.7%
	Romance	French	57	1	0	1	98.2%
		Italian	52	1	1	1	84.6%
		Portuguese	61	1	0	1	77.0%

		Romanian	31	1	1	1	74.2%
		Spanish	36	1	1	1	80.6%
	Slavic	Polish	53	1	1	1	88.7%
		Russian	47	1	1	1	87.2%
Japanese	Japanese	Japanese	55	0	0	1	92.7%
Kartvelian	Kartvelian	Georgian	15	1	0	1	80.0%
Korean	Korean	Korean	22	0	0	1	90.9%
Sino-Tibetan	Chinese	Mandarin Chinese	46	0	0	1	69.6%
Tai-Kadai	Kam-Tai	Thai	20	1	0	1	80.0%
Turkic	Turkic	Turkish	37	1	0	1	81.1%
Uralic	Finnic	Estonian	43	1	1	1	100.0%
		Finnish	18	1	1	1	100.0%
	Ugric	Hungarian	34	1	1	1	94.1%
Total N/Percentage of occurrence			903	84%	44%	100%	87.3%

TABLE II. Counts of participants per language and language family in the field experiment. The table is ordered alphabetically by language, within family and genus. The participant sample is discussed in Section A; the rhotic coding is discussed in Section D; the “match” variable is discussed in Section E.

Family	Genus	Name	N of participants	r/l contrast	[ɾ] as main r-sound	[ɾ] as allophone	“match”
Afro-Asiatic	Amazigh	Tashlhiyt	20	1	1	1	100.0%

Arawakan	Eastern Arawakan	Palikúr	8	0	0	0	100.0%
Austronesian	Oceanic	Daakie	12	1	1	1	100.0%
Indo-European	Germanic	English (UK)	55	1	0	1	98.2%
		German	19	1	0	1	94.7%
	Romance	Brazilian Portuguese	13	1	0	1	92.3%
Total N/Percentage of occurrence			127	83%	33%	83%	97.6%

In the web experiment, 143 participants (16%) lacked an r/l contrast in their first language, with only one of the participants also not using the r/l contrast in any second language. A total of 372 participants (41.2%) spoke a first language that uses the alveolar trill [r] as the primary r-sound; 531 participants (41.2%) spoke a first language for which the trill was not the primary variant. A total of 295 participants (32.7%) spoke at least one second language that uses the alveolar trill [r] as the primary r-sound, as opposed to 486 participants (53.8%) with second language(s) in which the alveolar trill was the primary variant (122 participants did not speak any second language, 13.5%).

For the field experiment, only 8 participants (6% of the sample) spoke a first language that lacks an r/l contrast; all of these 8 participants also knew a foreign language that distinguishes phonemically between /r/ and /l/, which suggests that the entire sample knew at least one language that feature an r/l contrast. A total of 32 participants (25.2%) spoke at least one language natively in which the alveolar trill [r] was the primary r-sound, as opposed to 95 participants (74.8%) who spoke native languages where this was not the case. A total of 46 participants (36.2%) reported speaking at least one second language in which the alveolar trill [r] was the primary r-sound; 28 participants (22.0%) spoke a second language without an alveolar trill as the primary variant; 53 participants (41.7%) reported speaking no second language.

As can be seen across both, Table I and Table II, participants from almost all languages, except for Palikúr, spoke at least one language in which trilled [r] could feature as an allophone. Here, our definition of allophones is intentionally broad, encompassing any variant of a phoneme that may appear in specific contexts or as free variation. This includes cases where [r] might be considered an non-standard or a less common variant. For example, while the r-sound of standard German is not an alveolar trill, it does feature in certain dialects and is traditionally also associated with singing and theatre performances (called *Bühnen-r* "stage r" by Theodor Siebs). Similarly, although [r] is not the main allophone in French, it is retained by some speakers and in certain regional varieties. Likewise, while Japanese is not typically known for having trilled [r] as a primary or standard allophone, this sound can occur in certain forms of speech, such as "gangster speech" (Sreetharan, 2004, p. 97). Also in American English, which does not have trilled [r] as part of its standard phonemic inventory, one can find instances thereof in comedic displays, or advertisements, where it is used for expressive purposes (Winter et al., 2022, p. 5). To establish that trilled [r] can occur as an allophone, we collected data systematically through published literature, and, where necessary, through online sources or direct recordings, without any a priori assumptions about what to expect from each language. When a language is coded as "[r] as allophone" but not "[r] as the main sound," this implies that the trilled [r] is less frequent in those languages, as it appears in specific contexts rather than being a primary feature of the language's phonological system. However, its presence as an allophone indicates that it is still embedded within the language's phonology, albeit in a more limited and context-dependent manner.

E. **Statistical analysis**

Throughout all analyses, we use R (R Core Team, 2019) together with the tidyverse package (Wickham et al., 2019) for data processing and visualization. All statistical models are a version of multilevel Bayesian logistic regression implemented in brms (Bürkner, 2017). In both the web experiment and the field experiment, each participant contributed two data points. We collapsed both data points into a single data point per participant, a variable we call “match,” and the main dependent variable of our logistic regression models. For this variable, we only coded cases as match (1) when they were complete matches, i.e., a participant matched the jagged line to [r] *and* they matched the flat line to [l]. Complete mismatches (matching [l] to the jagged line and [r] to the flat line), as well as partial matches (e.g., matching [r] to both the jagged and the flat line) were both coded as mismatch (0) (cf. Ćwiek et al., 2022). If we assume that both responses are independent, chance for the match variable would be at 25%. However, it is likely that the second response is influenced by the first one, in which case chance would exceed 25%, and would be 50% if the second response was entirely locked to the first. Especially because for the field experiment, both sound files were presented first, we took a conservative approach by assuming complete dependence between the responses and chose 50% as our chance level baseline to measure matching performance.

The first model we report is a logistic regression model that includes two fixed effects: an intercept, and a fixed effect for “experiment,” which is a treatment-coded indicator variable representing the difference between the web experiment (0 = reference level) and the field experiment (1). This model includes random intercepts for language, family, and Autotyp area, defined by Nichols et al. (2013). The Autotyp areas are geographic regions grouping languages based on shared linguistic features and historical interactions, rather than genetic relationships. We then assess the impact of the language-level predictors in line with our rhotic coding as described in Section D, with one model testing the fixed effects “has trilled [r] in L1” and “has trilled [r] in L2,” and another model testing the fixed effects “has r/l contrast in L1” and “has r/l contrast in L2.” These variables were treatment-coded, with not having [r] or not having an r/l contrast as the reference level (= 0). We fitted separate models for these two types of predictors because data for the r/l contrast variable was heavily unbalanced, with very few languages not making this contrast. We did not fit a model for the "[r] as allophone" variable because as Tables I and II show, there is not enough variation between languages to test the impact of this factor.

All models included the same random intercepts as described above. Random slopes for the rhotic predictors were impossible to implement as there was generally no variation for these predictors within language family or Autotyp area (cf. Table I and II). The only random slope that was possible to implement due to having enough variation within random effects levels was "has trilled [r] in L2" for language family, which we added to the model testing for these fixed effects. As "presentation order" was only controlled for in the web experiment, we tested this variable in a separate model fitted to data from the web experiment only (with by-language, by-family, and by-Autotyp area random slopes for order). As this predictor was roughly balanced (467 participants in the web experiment heard [r] first, 436 heard [l] first; 51.7% versus 48.2%), we sum-coded this predictor (-1 = [r] first, +1 = [l] first) to aid the interpretation of the intercept, which then represents the grand average matching probability.

We used Student-t distributed priors for the intercept (degrees of freedom = 3, scale = 2.5) and random effect standard deviations. We also used Student-t distributed priors (degrees of freedom = 5, scale = 2.5) for all fixed effects slopes. We used LKJ(2) priors for all random effect correlation terms. Prior predictive simulations showed that these priors accommodate our data well. We additionally verified that our fitted models adequately captured plausible data-generating processes via posterior predictive simulations. All models were estimated using Markov Chain Monte Carlo simulation with four chains à 10,000 iterations (4,000 warm-up samples excluded, thin = 2 to reduce disk space for fitted models), which resulted in 12,000 posterior samples used for inference.

We list descriptive percentages for "match" in Table I and II. The estimates of individual languages stemming from the statistical model seen in Fig. 2 differ from the descriptive values due to shrinkage: in multilevel models, information from the group level results is used to inform individual random effects estimates, which are drawn towards the mean.

III. RESULTS

On average, matching probability was very high, with the descriptive mean lying at 88.5% across both the web and the field experiments. Average matching was high for speakers from all languages in the sample, with the highest being 100% for Estonian and Finnish speakers, and the lowest being 70% for Albanian and Mandarin Chinese speakers.

The multilevel logistic regression coefficient estimates the average matching for web experiment as 88.2%, with a 95% credible interval (CrI) of [81.7%, 92.9%]. For the field experiment, the posterior mean is 97.5%, 95% CrI [92.9%, 99.2%]. The credible intervals for both experiments are far above the chance threshold, with the posterior probability of exceeding chance being $p(>50\%)=1.0$ for both samples. This indicates that given this data, model, and priors, we can be very certain that the cross-linguistic average in both samples exceeds chance. The slope of the fixed effect of "experiment" was positive, indicating higher average matching for the field experiment than the web experiment (logit estimate = +1.64, SE = 0.68, 95% CrI [0.57, 2.79]). The posterior probability of this coefficient having the same sign was $p(\beta>0)=0.99$, indicating high certainty that matching was higher in the field experiment than the web experiment. Fig. 2 shows the posterior estimates and 95% credible intervals for all languages sorted by average matching, with the box highlighting that results from the field languages all have the highest averages.

With respect to the rhotic predictors, descriptive statistics indicate that the participants whose native language have the alveolar trill [r] as the primary rhotic variant have a slightly lower proportion of matches (86.6%) than those without (89.8%). This difference, although small in terms of effect size, is indicated to be quite certain given this model, data and priors: the posterior probability of this coefficient having the same sign was high $p(\beta<0)=0.99$ (logit coefficient: -0.93, *SE*

= 0.4, 95% CrI [-1.6, -0.3]). This effect can be seen in Fig. 2, where languages within which the alveolar trill [r] is the main allophone (color: orange) appear relatively more towards the left of the plot, compared to the other languages (color: blue). There was no such effect for speaking a second language with an alveolar trill [r] (logit estimate = +0.5, *SE* = 0.93, 95% CrI: [-0.8, +2.2]), with the posterior probability of being positive at *p*(β>0)=0.71, indicating that this specific result is bound up with considerable uncertainty. Similarly, results from the model including the predictors for whether r/l are phonologically distinguished in the language were inconclusive (coefficient for r/l in L1: -0.29, *SE* = 0.79, 95% CrI: [-1.0, +1.6], *p*(β>0)=0.66; r/l in L2: -1.33, *SE* = 2.5, 95% CrI: [-5.8, +2.0], *p*(β>0)=0.69). This is also apparent when looking at the plot in Fig. 2, where circles indicate languages without an r/l contrast, which appear amongst those languages with the highest average matching (field experiment: Palikúr), as well as amongst those languages with the lowest average matching (web experiment: Mandarin Chinese), and everything in between.

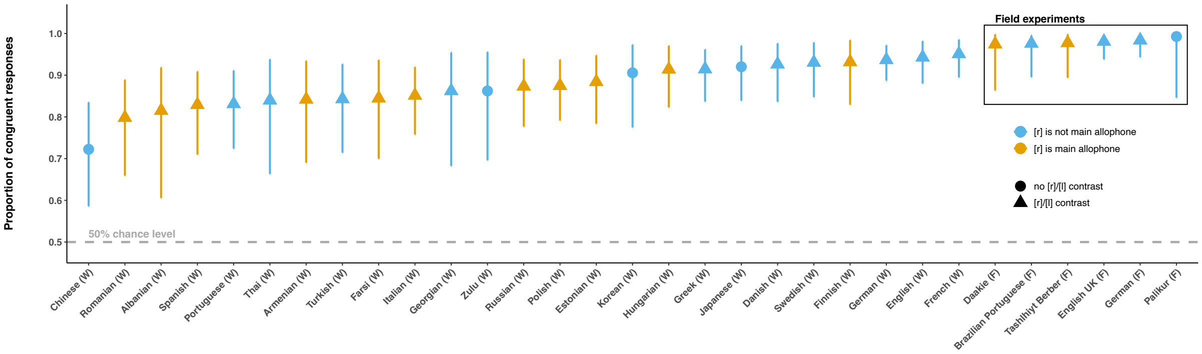

FIG 2. Posterior means (triangles: r/l contrast, circles: no r/l contrast) for results from each language, with colors representing whether [r] is the main allophone (orange) or not (blue); the dashed line indicates our conservative 50% chance level, highlighting that 95% credible intervals (line segments) are far above this threshold; as discussed in Section E individual estimates (posterior

means) presented differ from the descriptive accuracies due to shrinkage, which draws them towards the group means.

As discussed in the methods section, the effect of order ([r] played first versus [l] played first) was controlled only for the web experiment. When looking at the first trial only, the jagged line was chosen 94.2% of the time when [r] was played first, and the flat line was chosen 83.7% of the time when [l] was played first. This indicates that both [r] and [l] alone are matched correctly, but [r] more consistently so. The model with an effect of order fitted to the subset of the data only from the web experiment indicates that the difference between [r]-first and [l]-first trials is relatively certain (logit estimate for [l] first: -1.0, *SE* = 0.6, 95% CrI [-2.1, -0.1]), with a high posterior probability of being of the same sign, $p(\beta>0)=0.96$.

II. **DISCUSSION**

Typological analyses of spoken vocabularies have found a statistical bias towards the occurrence of /r/ in words that refer to rough qualities of texture (Winter et al., 2022). The source of this bias has been hypothesized to be an iconic correspondence between roughness and the acoustic/articulatory properties of the alveolar trill [r] in particular (although see Anselme et al., 2022), but experimental evidence for this connection was lacking. Here, to investigate the basis for the correspondence between r-sounds and rough meanings, we conducted an experiment to test whether speakers of different languages associate an alveolar trill [r] with a jagged/rough line, and in contrast, an alveolar lateral approximant [l] with a flat/smooth line, in a task that emphasizes touch as much as possible by asking participants to imagine moving their finger across each line. In two experiments, one online and the other on-site, participants – including speakers of 28 different languages from 10 language families – listened to recordings of an [r] and an [l], and were asked to match each sound to an image of either a jagged line or a straight line.

We found a strong effect overall: participants matched [r] with the jagged line/rough surface and [l] with the smooth line/smooth surface an estimated 88% of trials for the online experiment and 98% of trials for the field experiment, well above the conservative baseline level of 50%. It is noteworthy that this matching probability is about 15% higher than what was observed for the *bouba/kiki* effect in a study using the same sample of speakers and a highly comparable experimental design that also involved sequential matching (Ćwiek et al., 2022). Moreover, in stark contrast to our current experiment, the *bouba/kiki* effect was found to have exceptions among language groups, with some groups not showing the effect. In the present data, *all* of the language groups in our sample showed the effect, i.e., the pattern is exceptionless, with each group showing a matching probability that is well above chance. These results indicate that the [r]/[l] crossmodal correspondence is extremely strong and one of the most cross-culturally robust cases of sound symbolism documented to date.

There are several other results worth highlighting. First, past research on the *bouba/kiki* effect has shown that tasks involving paired matching greatly amplify the effect (see discussion in Nielsen and Rendall, 2011, 2012). We believe this is the most likely explanation for why in the present study, the field experiment showed overall higher matching than the web experiment, by about 10%. In the web experiment, each response followed each auditory stimulus (see Section C). In the field experiment, people gave their two responses only after hearing both sounds, thus facilitating paired matching.

Another notable result was the order effect observed in the web experiment, such that on first trials, [r] was matched to the jagged line more consistently than [l] was matched to the flat line, by about 10%. The fact that both these percentages were well bove 50% for first trials indicates that both [r] and [l] independently carried strong iconic associations with their respective lines/textures, even as the effect was somewhat stronger for [r]. This highlights the advantage of sequential matching, which allows teasing apart the relative contribution of each stimulus, in contrast to paired matching for which it is unknown how much each stimulus contributes to the overall picture. Notably, previous studies using a sequential matching design found a similar pattern with the *bouba*/*kiki* effect, where the *bouba* stimulus is more consistently associated with the round shape, than *kiki* with the angular shape (Ćwiek et al., 2022; Fort et al., 2018; Margiotoudi et al., 2019; Yang et al., 2019).

Importantly, the effect we observed here is clearly present for speakers of languages with differing r-sounds and phoneme inventories with respect to these sounds: matching exceeds chance regardless of whether speakers spoke a first or second language in which [r] was the primary realization or not, and regardless of whether they spoke a language that phonologically distinguished between /r/ and /l/ sounds. Even though matching was high regardless of the phonological and phonetic characteristics of rhotics in speakers' first and second languages, we found a small but reliable effect where matching was reduced for speakers of languages in which trilled [r] is the primary variant. One possible explanation for this result is that when this sound is used as a contrastive phoneme within a language, and therefore regularly serves the phonemic function of distinguishing arbitrary words, its iconic associations may be reduced. This suggests that the extent to which a sound triggers iconic associations is malleable, and modulated by the degree to which a sound is embedded within the phonological grammar of a language.

This result may be reflected in historical situations in which languages come to acquire an alveolar trill as part of their standard phonemic inventory through contact with other languages. For example, Campbell (2004, p. 68) discusses a scenario where speakers of two Mayan languages, Chol and Tzotzil, had no trilled [r] sound before exposure to Spanish. After the sound was introduced into both of these languages via loan words, this new foreign sound, "which apparently seemed exotic to the speakers of these Mayan languages" (p. 68), came first to be employed specifically in onomatopoeias and expressive vocabulary and only later ventured into the general lexicon where it featured in arbitrary contrasts. This historical case suggests that for speakers of languages that do not already have an alveolar trill within their native phoneme inventory, this sound initially carries high expressive potential before it becomes more embedded within the grammar. However, more work is needed to ascertain whether it is specifically the conventional use of the alveolar trill [r] that is driving the weakening of the effect in our study. One way of testing this hypothesis more directly would be to quantify the functional load carried by /r/ sounds in different languages, e.g., in terms of how many meanings are distinguished by /r/ in the lexicon (cf. Wedel et al., 2013a, 2013b). Our current results would predict that the more meanings depend on /r/, that is, the higher its functional load, the more its expressive associations should be diminished.

Our results also speak to a long-standing debate in sound symbolism research: whether the analogies underpinning the iconicity of speech sounds are primarily rooted in acoustic or articulatory factors (e.g., Diffloth, 1994; Margiotoudi et al., 2019; Sapir, 1929; Sidhu and Vigliocco, 2022; Thompson and Do, 2019; Vainio and Vainio, 2021). Are iconic correspondences based in the acoustics of speech sounds, or are they based on articulatory factors, including proprioception – how it feels to articulate the sounds with the vocal tract – and vision – the visible features of the mouth and face involved in articulating the sounds? For the *bouba*/*kiki* effect, it has been found that resemblances based on acoustic factors are sufficient to carry the effect, as it also occurs with reversed speech that cannot be articulated (Passi and Arun, 2022), as well as in stimuli that are filtered to be non-speech objects to listeners (Silva and Bellini-Leite, 2019). Moreover, the effect is not modulated by seeing videos of speakers pronouncing the nonce words *bouba* and *kiki,* and if anything, weakened by viewing such articulations (Sidhu and Vigliocco, 2022).

The alevolar trill is interesting from this perspective, as it is a sound that is notoriously difficult to articulate, requiring precise articulatory and aerodynamic control (Solé, 2002). To achieve the distinctive mode of trilled tongue movement, speakers must "position the tongue and apply the correct amount of pressure against the alveolar ridge" to allow pressure to "overcome occlusion while maintaining ability for the tongue to recoil" (Olsen, 2016, p. 317). Evidence from first and second language acquisition shows that alveolar trills are acquired late (Ball et al., 2001; Jiménez, 1987; Kehoe, 2018; Mendoza, 2000), and indeed, some native speakers never learn to articulate the sound (Solé, 2002, p. 656) – an outcome that is common enough to receive a label in some languages, such as Italian *erre moscia* "weak r," used to refer to Italian speakers, including native speakers, who cannot master trills. From this perspective, it is interesting that speakers of Palikúr, the only language in which trilled [r] never occurs as an allophone, performed matching at ceiling (100%). Together with the evidence from languages in which the trill is not regularly used, such as Mandarin Chinese and Japanese, this shows that even when speakers cannot produce trilled [r], they still perceive the sound to be more fitting for the jagged rather than the flat line. Indeed, models such of the acquisition of non-native consonants, such as PAM (Perceptual Assimilation Model, Best et al., 2001), suggest that non-native sounds that cannot be assimilated to any existing sounds in a language may be perceived as non-speech sounds, which essentially are mere acoustic objects without learned articulatory representations. Thus, similar to what has been observed for *bouba/kiki* (Passi and Arun, 2022; Silva and Bellini-Leite, 2019), this suggests that the acoustics of the alveolar trill [r] alone are sufficient to carry the effect. Future work performing acoustic manipulations of [r] sounds similar to those that have been conducted for *bouba/kiki* (Passi and Arun, 2022; Silva and Bellini-Leite, 2019) could be used to lend further support for this interpretation of our results.

In using only one /r/ and one /l/ stimulus, our study was not explicitly set up to experimentally manipulate acoustic factors and test what specific cues drive the matching we observed. That being said, a very likely cognitive mechanism that explains our results is the fact that independent of speech, people associate spatial frequencies crossmodally with the frequency of amplitude modulation (Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013; Sherman et al., 2013). Our stimuli differ exactly in these two characteristics, albeit in a categorical manner: one stimulus has spatial frequency, the other one does not. And one sound involves repeated closure (and hence cyclical amplitude modulation), the other one does not. Given that prior research in perceptual psychology has shown a correspondence between the same visual and auditory feature that are also contrastive in our study, we think that this is the most likely mechanism. Interestingly, amplitude modulation also turns out to be an important cue for the *bouba/kiki* effect (Fort and Schwartz, 2022). It has to be borne in mind, however, that Anselme et al. (2022) have recoded the phonetic characteristics of r-sounds in Winter et al.'s lexical data, which suggested that all r-like sounds may be associated with roughness in texture vocabularies. This suggests that there may be other aspects to the perceived roughness of r-sounds, on top of the amplitude modulation that differed saliently in this task.

Finally, another point for future research relates to our use of visual images to represent rough and smooth textures, with instructions for participants to imagine the feel as they move their finger along the lines. Our study aimed to shed experimental light on the source of the lexical patterns found by Winter et al. (2022) related to words describing rough textures, and yet, our evidence is somewhat indirect, mediated through the use of visual images. In this respect, it is interesting to note the deep similarity between roughness and jaggedness, which can be seen as related multisensory properties that vary in spatial frequency. For comparison to the effect found here, the *bouba/kiki* effect also exhibits a strong tactile component, having also been obtained with felt rather than seen shapes (Ciaramitaro et al., 2021; Fryer et al., 2014; Graven and Desebrock, 2018; Sakamoto and Watanabe, 2018). And evidence obtained with Italian speakers shows that *bouba/kiki*-type words are not only matched to shapes, but also to surfaces differing in roughness (Etzi et al., 2016). Indeed, the crossmodal correspondence between spatial frequency and amplitude modulation also works between vibrotactile frequency and amplitude modulation (Guzman-Martinez et al., 2012), suggesting that the same feature — spatial frequency — matters for both modalities. Thus, even though our experimental stimuli are ambiguous with respect to vision/touch, the dimension of shape we investigate is conceptually similar to, and associated with, textural roughness. The fact that *bouba/kiki* effects work in both vision and touch, including with stimuli differing in roughness only, suggest that our results might also carry over to an experimental design that involved a genuine touch component, which was not possible in our web-based experiment. Future work can use textural stimuli to further hone in on the connection between speech sounds and touch alone.

To conclude, we found – in a large cross-linguistic experiment spanning a diverse sample of participants speaking 28 languages from 12 different language families, including participants from cultures with little access to technology and globalized culture – that trilled [r] was overwhelmingly associated with a jagged/rough line, and correspondingly, [l] with a flat/smooth line. While the average effect was always found regardless of the phonetic and phonological characterics of rhotics in participants' respective languages, it was somewhat weakened for speakers who use trilled [r] as the primary variant, suggesting the conventional use of this sound as a phoneme may diminish its iconic power. Nevertheless, the effect was extremely strong, even stronger than what has been observed for the widely studied *bouba/kiki* effect. In contrast to the *bouba/kiki* effect, which is not obtained for speakers from all languages (Ćwiek et al., 2022; Styles and Gawne, 2017), the r/l effect observed here was obtained without exception for all language groups in our sample, suggesting it may be one of the most cross-linguistically robust cases of sound symbolism documented to date.

SUPPLEMENTARY MATERIAL

The supplementary material, including the stimuli, the data, and the analysis scripts, is available in the OSF repository: https://osf.io/mjcnq/

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation, grant number FU 791/6-1. The authors wish to thank the co-authors of the main study: Christoph Draxler, Eva Liina Asu, Katri Hiovain, Sofia Koutalidis, Manfred Krifka, Pärtel Lippus, Gary Lupyan, Nathalie Schümchen, Ádám Szalontai, and Özlem Ünal-Logacev. We also thank Mohammad Ali Nazari, Samer Al Moubayed, Anna Ayrapetyan, Carla Bombi Ferrer, Nataliya Bryhadyr, Chiara Celata, Ioana Chitoran, Taehong Cho, Soledad Dominguez, Cornelia Ebert, Mattias Heldner, Mariam Heller, Louis Jesus, Enkeleida Kapia, Soung-U Kim, James Kirby, Jorge Lucero, Konstantina Margiotoudi, Mariam Matiashvili, Feresteh Modaressi, Scott Moisik, Oliver Niebuhr, Catherine Pelachaud, Zacharia Pourtskhvanidze, Pilar Prieto, Vikram Ramanarayanan, Oksana Rasskazova, Daniel Recasens, Amélie Rochet-Capellan, Mariam Rukhadze, Johanna Schelhaas, Vera Scholvin, Frank Seifart, Stavros Skopeteas, SOS children's village Armenia, Katarzyna Stoltmann, and Martti Vainio for being involved in the translation or distribution of the survey. We thank Timo Roettger for supplying the plot theme function. Finally, we thank Catherine Best for her advice on the perception of the alveolar trill.

AUTHORS’ CONTRIBUTIONS

AĆ, SF, MP, and BW conceived and designed the study. AĆ, DD, SF, SK, GO, JP, MP, CP, SR, RR, JZ, and BW translated (or arranged translation of) and distributed the surveys. AĆ and RA collected the data on r-variants and performed the rhotic coding. AĆ, DD, and BW performed the statistical analyses. AĆ wrote the first draft. MP and BW revised the manuscript. AĆ, DD, SF, SK, MP, and BW edited revisions of the manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The ethics approval has been granted to the PSIMS project by the German Linguistics Society Ethics Commission with the number #2018-02-180912.

DATA AVAILABILITY

All data, scripts, and models are available in the OSF repository: https://osf.io/mjcnq/

REFERENCES (BIBLIOGRAPHIC STYLE)

Anselme, R., Pellegrino, F., and Dediu, D. (**2022**). "R You Sure That Your /r/ Is Trilled? A Methodological Caveat," Available: https://cnrs.hal.science/hal-03900020, (date last viewed: 09-Aug-24). Presented at the Joint Conference on Language Evolution. Retrieved from https://cnrs.hal.science/hal-03900020

Anselme, R., Pellegrino, F., and Dediu, D. (**2023**). "What's in the r? A review of the usage of the r symbol in the Illustrations of the IPA," Journal of the International Phonetic Association, **53**, 1003–1032. doi:10.1017/S0025100322000238

Aryani, A., Isbilen, E. S., and Christiansen, M. H. (**2020**). "Affective Arousal Links Sound to Meaning," Psychol Sci, **31**, 978–986. doi:10.1177/0956797620927967

Auracher, J. (**2017**). "Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance," PLOS ONE, **12**, e0187196. doi:10.1371/journal.pone.0187196

Ball, M. J., Müller, N., and Munro, S. (**2001**). "The acquisition of the rhotic consonants by Welsh-English bilingual children," International Journal of Bilingualism, **5**, 71–86. doi:10.1177/13670069010050010401

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," The Journal of the Acoustical Society of America, **109**, 775–794. doi:10.1121/1.1332378

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., and Majid, A. (**2022**). "Over-reliance on English hinders cognitive science," Trends in Cognitive Sciences, **26**, 1153–1170. doi:10.1016/j.tics.2022.09.015

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (**2016**). "Sound–meaning association biases evidenced across thousands of languages," Proceedings of the National Academy of Sciences, **113**, 10818–10823. doi:10.1073/pnas.1605782113

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., and 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.

Bürkner, P.-C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," Journal of Statistical Software, **80**, 1–28.

Campbell, L. (**2004**). *Historical linguistics*, MIT Press, Cambridge, MA.

Ciaramitaro, V., Kelly, J., and Nguyen, C. (**2021**). "The feeling of 'kiki': select tactile exposure or visual imagery can enhance abstract audio-tactile crossmodal correspondences, the bouba/kiki effect, early in development," Journal of Vision, **21**, 2052. doi:10.1167/jov.21.9.2052

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., et al. (**2021**). "Novel vocalizations are understood across cultures," Scientific Reports, **11**, 10108. doi:10.1038/s41598-021-89445-4

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., et al. (**2022**). "The bouba/kiki effect is robust across cultures and writing systems," Philosophical

Transactions of the Royal Society B: Biological Sciences, **377**, 20200390. doi:10.1098/rstb.2020.0390

Diffloth, G. (**1994**). "i: big, a: small," In J. J. Ohala, L. Hinton, and J. Nichols (Eds.), Sound symbolism, Cambridge University Press, Cambridge, UK, pp. 107–113.

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**, 603–615.

Draxler, C. (**2011**). "Percy-an HTML5 framework for media rich web experiments on mobile devices," INTERSPEECH 2011: Twelfth Annual Conference of the International Speech Communication Association, ISCA, Florence, Italy, 3339–3340. Retrieved from https://www.isca-archive.org/interspeech_2011/draxler11b_interspeech.pdf

Etzi, R., Spence, C., Zampini, M., and Gallace, A. (**2016**). "When sandpaper is 'Kiki'and satin is 'Bouba': an exploration of the associations between words, emotional states, and the tactile attributes of everyday materials," Multisensory Research, **29**, 133–155.

Fitch, W. T. (**1994**). *Vocal tract length perception and the evolution of language* (BA Thesis), Brown University, Providence.

Fónagy, I. (**1961**). "Communication in poetry," Word, **17**, 194–218.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., and Tsuji, S. (**2018**). "Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition," Developmental Science, **21**, e12659. doi:10.1111/desc.12659

Fort, M., and Schwartz, J.-L. (**2022**). "Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects," Sci Rep, **12**, 19172. doi:10.1038/s41598-022-23623-w

Fryer, L., Freeman, J., and Pring, L. (**2014**). "Touching words is not enough: How visual experience influences haptic–auditory associations in the 'Bouba–Kiki' effect," Cognition, **132**, 164–173.

Graven, T., and Desebrock, C. (**2018**). "Bouba or kiki with and without vision: Shape-audio regularities and mental images," Acta Psychologica, **188**, 200–212. doi:10.1016/j.actpsy.2018.05.011

Greenberg, J. H., and Jenkins, J. J. (**1966**). "Studies in the psychological correlates of the sound system of American English," Word, **22**, 207–242.

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

Hammarström, H., Forkel, R., Haspelmath, M., and Bank, S. (**2020**). "glottolog/glottolog: Glottolog database 4.3.," doi:10.5281/zenodo.4061162

Haynie, H., Bowern, C., and LaPalombara, H. (**2014**). "Sound symbolism in the languages of Australia," PLoS One, **9**, 1–16. doi:10.1371/journal.pone.0092852

Henrich, J., Heine, S. J., and Norenzayan, A. (**2010**). "Most people are not WEIRD," Nature, **466**, 29–29. doi:10.1038/466029a

Hollins, M., and Bensmaïa, S. J. (**2007**). "The coding of roughness," Canadian Journal of Experimental Psychology, **61**, 184. doi:10.1037/cjep2007020

Hoshi, H., Kwon, N., Akita, K., and Auracher, J. (**2019**). "Semantic associations dominate over perceptual associations in vowel–size iconicity," i-Perception, **10**, 2041669519861981. doi:10.1177/2041669519861981

Huang, Y.-H., Pratoomraj, S., and Johnson, R. C. (**1969**). "Universal magnitude symbolism," Journal of Verbal Learning & Verbal Behavior, **8**, 155–156. doi:10.1016/S0022-5371(69)80028-9

Jiménez, B. C. (**1987**). "Acquisition of Spanish consonants in children aged 3–5 years, 7 months," Language, Speech, and Hearing Services in Schools, **18**, 357–363. doi:10.1044/0161-1461.1804.357

Johansson, N., Anikin, A., Carling, G., and Holmer, A. (**2019**). "The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features," Linguistic Typology, **24**, 253–310.

Johnson, R. C. (**1967**). "Magnitude symbolism of English words," Journal of Verbal Learning and Verbal Behavior, **6**, 508–511. doi:10.1016/S0022-5371(67)80008-2

Kehoe, M. (**2018**). "The development of rhotics: a comparison of monolingual and bilingual children," Bilingualism: Language and Cognition, **21**, 710–731. doi:10.1017/S1366728916001279

Knoeferle, K., Li, J., Maggioni, E., and Spence, C. (**2017**). "What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings," Scientific Reports, **7**, 5562. doi:10.1038/s41598-017-05965-y

Köhler, W. (**1929**). *Gestalt psychology*, Liveright, New York, NY.

Lederman, S. J. (**1974**). "Tactile roughness of grooved surfaces: The touching process and effects of macro-and microsurface structure," Perception & Psychophysics, **16**, 385–395. doi:10.3758/BF03203958

Lederman, S. J. (**1979**). "Auditory texture perception," Perception, **8**, 93–103. doi:10.1068/p080093

Lederman, S. J. (**1983**). "Tactual roughness perception: spatial and temporal determinants," Canadian Journal of Psychology/Revue canadienne de psychologie, **37**, 498–511.

Lederman, S. J., and Abbott, S. G. (**1981**). "Texture perception: Studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics," Journal of

Experimental Psychology: Human Perception and Performance, **7**, 902–915. doi:10.1037/0096-1523.7.4.902

Levickij, V. V. (**2013**). "Phonetic symbolism in natural languages," Glottotheory, **4**, 72–91. doi:10.1524/glot.2013.0006

Lockwood, G., and Dingemanse, M. (**2015**). "Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism," Front Psychol, **6**, 1246. doi:10.3389/fpsyg.2015.01246

Margiotoudi, K., Allritz, M., Bohn, M., and Pulvermüller, F. (**2019**). "Sound symbolic congruency detection in humans but not in great apes," Scientific Reports, **9**, 1–12. doi:10.1038/s41598-019-49101-4

Marks, L. E. (**1978**). *The unity of the senses: Interrelations among the modalities*, Academic Press, New York.

Mendoza, G. C., Elvira (**2000**). "Acoustic characteristics of trill productions by groups of Spanish children," Clinical Linguistics & Phonetics, **14**, 587–601. doi:10.1080/026992000750048125

Moran, S., McCloy, D., and Wright, R. (**2014**). "PHOIBLE online 2.0," Retrieved from https://phoible.org/. Retrieved from https://phoible.org/

Newman, S. S. (**1933**). "Further experiments in phonetic symbolism," The American Journal of Psychology, **45**, 53–75.

Nielsen, A., and Rendall, D. (**2011**). "The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon.," Canadian Journal of Experimental Psychology, **65**, 115–124. doi:10.1037/a0022268

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

Ohala, J. J. (**1983**). “Cross-language use of pitch: an ethological view,” Phonetica, **40**, 1–18. doi:10.1159/000261678

Olsen, M. K. (**2016**). “Limitations of the influence of English phonetics and phonology on L2 Spanish rhotics,” Borealis – An International Journal of Hispanic Linguistics, **5**, 313–331. doi:10.7557/1.5.2.3898

Orchard-Mills, E., Van der Burg, E., and Alais, D. (**2013**). “Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies,” Journal of Vision, **13**, 6. doi:10.1167/13.3.6

Parise, C., and Spence, C. (**2012**). “Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test,” Experimental Brain Research, **220**, 319–333. doi:10.1007/s00221-012-3140-6

Passi, A., and Arun, S. P. (**2022**). “The Bouba–Kiki effect is predicted by sound properties but not speech properties,” Attention, Perception, & Psychophysics, , doi: 10.3758/s13414-022-02619-8. doi:10.3758/s13414-022-02619-8

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**, 227.

R Core Team (**2019**). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

Ramachandran, V. S., and Hubbard, E. M. (**2001**). “Synaesthesia–a window into perception, thought and language,” Journal of Consciousness Studies, **8**, 3–34.

Sakamoto, M., and Watanabe, J. (**2018**). “Bouba/kiki in touch: Associations between tactile perceptual qualities and Japanese phonemes,” Frontiers in Psychology, **9**, 1–12. doi:10.3389/fpsyg.2018.00295

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

Sherman, A., Grabowecky, M., and Suzuki, S. (**2013**). “Auditory rhythms are systemically associated with spatial-frequency and density information in visual scenes,” Psychon Bull Rev, **20**, 740–746. doi:10.3758/s13423-013-0399-y

Sidhu, D. M., and Vigliocco, G. (**2022**). “I don’t see what you’re saying: The maluma/takete effect does not depend on the visual appearance of phonemes as they are articulated,” Psychon Bull Rev, , doi: 10.3758/s13423-022-02224-8. doi:10.3758/s13423-022-02224-8

Sidhu, D. M., Westbury, C., Hollis, G., and Pexman, P. M. (**2021**). “Sound symbolism shapes the English language: The maluma/takete effect in English nouns,” Psychonomic Bulletin & Review, , doi: 10.3758/s13423-021-01883-3. doi:10.3758/s13423-021-01883-3

Silva, D. M. R., and Bellini-Leite, S. C. (**2019**). “Cross-modal correspondences in sine wave: Speech versus non-speech modes,” Attention, Perception, & Psychophysics, **82**, 944–953. doi:10.3758/s13414-019-01835-z

Solé, M.-J. (**2002**). “Aerodynamic characteristics of trills and phonological patterning,” Journal of Phonetics, **30**, 655–688. doi:10.1006/jpho.2002.0179

Sreetharan, C. S. (**2004**). “Students, sarariiman (pl.), and seniors: Japanese men’s use of 'manly' speech register,” Language in Society, **33**, 81–107. doi:10.1017/S0047404504031045

Styles, S. J., and Gawne, L. (**2017**). “When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter,” i-Perception, **8**, 2041669517724807.

Tarte, R. D., and Barritt, L. S. (**1971**). “Phonetic symbolism in adult native speakers of English: Three studies,” Language and Speech, **14**, 158–168. doi:10.1177/002383097101400206

Thompson, A. L., and Do, Y. (**2019**). “Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones,” Glossa: a journal of general linguistics, **4**, 1–40. doi:10.5334/gjgl.872

Thorndike, E. L. (**1945**). "On Orr's hypotheses concerning the front and back vowels," British Journal of Psychology, **36**, 10.

Ultan, R. (**1978**). "Size-sound symbolism," In J. H. Greenberg (Ed.), Universals of human language, Stanford University Press, Stanford, CA, pp. 525–568.

Vainio, L., and Vainio, M. (**2021**). "Sound-action symbolism," Frontiers in Psychology, **12**, 1–13. doi:10.3389/fpsyg.2021.718700

Wedel, A., Jackson, S., and Kaplan, A. (**2013a**). "Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change," Lang Speech, **56**, 395–417. doi:10.1177/0023830913489096

Wedel, A., Kaplan, A., and Jackson, S. (**2013b**). "High functional load inhibits phonological contrast loss: A corpus study," Cognition, **128**, 179–186. doi:10.1016/j.cognition.2013.03.002

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., et al. (**2019**). "Welcome to the Tidyverse," Journal of Open Source Software, **4**, 1686. doi:10.21105/joss.01686

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., and Grawunder, S. (**2021**). "Rethinking the frequency code: a meta-analytic review of the role of acoustic body size in communicative phenomena," Philosophical Transactions of the Royal Society B: Biological Sciences, **376**, 20200400. doi:10.1098/rstb.2020.0400

Winter, B., and Perlman, M. (**2021**). "Size sound symbolism in the English lexicon," Glossa: a journal of general linguistics, **6**, 1–13. doi:10.5334/gjgl.1646

Winter, B., Sóskuthy, M., Perlman, M., and Dingemanse, M. (**2022**). "Trilled /r/ is associated with roughness, linking sound and touch across spoken languages," Scientific Reports, , doi: 10.1038/s41598-021-04311-7. doi:10.1038/s41598-021-04311-7

Yang, J., Asano, M., Kanazawa, S., Yamaguchi, M. K., and Imai, M. (**2019**). “Sound symbolism processing is lateralized to the right temporal region in the prelinguistic infant brain,” Scientific Reports, **9**, 1–10. doi:10.1038/s41598-019-49917-0