Vowel signatures in emotional interjections and nonlinguistic vocalizations expressing pain, disgust, and joy across languages[a] FREE

Maïa Ponsonnet ; Christophe Coupé ; François Pellegrino ; Aitana Garcia Arasco ;
Katarzyna Pisanski

Check for updates

View Online   Export Citation

Articles You May Be Interested In

The voices of anger and disgust: Acoustic correlates in three languages

J Acoust Soc Am (November 2006)

Kyoo, what an odd-sounding word! Acoustic analyses of a Cajun interjection

J Acoust Soc Am (March 2019)

Emotional speech recognition for tamil language using spatial feature fusion and SVM classifier

AIP Conf. Proc. (January 2024)

Vowel signatures in emotional interjections and nonlinguistic vocalizations expressing pain, disgust, and joy across languages[a]

Maïa Ponsonnet,[1,2] Christophe Coupé,[3] François Pellegrino,[1] Aitana Garcia Arasco,[1] and Katarzyna Pisanski[1,4,5,b]

[1]*Dynamique Du Langage, CNRS et Université Lumière Lyon 2, Lyon, France*

[2]*School of Social Sciences, The University of Western Australia, Perth, Australia*

[3]*Department of Linguistics, The University of Hong Kong, Hong Kong SAR, China*

[4]*ENES Bioacoustics Research Laboratory, University Jean Monnet of Saint-Etienne, CRNL, CNRS, Saint-Etienne, France*

[5]*Institute of Psychology, University of Wrocław, Wrocław, Poland*

ABSTRACT:

In this comparative cross-linguistic study we test whether expressive interjections (words like *ouch* or *yay*) share similar vowel signatures across the world's languages, and whether these can be traced back to nonlinguistic vocalizations (like screams and cries) expressing the same emotions of pain, disgust, and joy. We analyze vowels in interjections from dictionaries of 131 languages (over 600 tokens) and compare these with nearly 500 vowels based on formant frequency measures from voice recordings of volitional nonlinguistic vocalizations. We show that across the globe, pain interjections feature a-like vowels and wide falling diphthongs ("ai" as in *Ayyy!* "aw" as in *Ouch!*), whereas disgust and joy interjections do not show robust vowel regularities that extend geographically. In nonlinguistic vocalizations, all emotions yield distinct vowel signatures: pain prompts open vowels such as [a], disgust schwa-like central vowels, and joy front vowels such as [i]. Our results show that pain is the only affective experience tested with a clear, robust vowel signature that is preserved between nonlinguistic vocalizations and interjections across languages. These results offer empirical evidence for iconicity in some expressive interjections. We consider potential mechanisms and origins, from evolutionary pressures and sound symbolism to colexification, proposing testable hypotheses for future research. © 2024 Acoustical Society of America.

https://doi.org/10.1121/10.0032454

I. INTRODUCTION

Across human cultures, people habitually vocalize when experiencing pain or emotional states such as joy and disgust. These vocal bursts may be entirely nonlinguistic, for example, cries of pain or amused bouts of laughter, but also frequently contain linguistic elements as in the case of expressive interjections such as *ouch*, *wow*, or *oops*. Are the forms of these interjections completely arbitrary, or do they share some common acoustic features across languages?

In this exploratory study, we examine the hypothesis that the forms of expressive interjections are not fully arbitrary, by testing whether interjections used to express distinct affective experiences such as pain, disgust, and joy share similar vowel patterns across diverse human languages. In a follow-up study, we further assess the possibility that the vowel patterns observed in these interjections may derive from their nonlinguistic counterparts, i.e., nonlinguistic vocalizations expressing the same emotions, through processes of conventionalization

or imitation. We argue that such processes could give rise to "iconicity" in expressive interjections, in the sense that their forms may be linked non-arbitrarily to their meanings (Perlman *et al.*, 2015).

A. Definitions, background, and hypotheses

Nonlinguistic vocalizations, sometimes referred to as nonverbal vocalizations (Pisanski *et al.*, 2022), or affect/vocal bursts (Brooks *et al.*, 2023; Cowen *et al.*, 2019; Scherer, 2019; Schröder, 2003), are here defined as vocal sounds that do not meaningfully resemble any word in the speaker's given language, such as a scream of joy or cry of pain (Pisanski *et al.*, 2022, for review). While researchers have been interested in human nonlinguistic vocalizations for more than a century (Darwin, 1872), research in this area is now expanding rapidly (Scherer, 2021). Inspired by research on nonhuman animal communication (e.g., Morton, 1977), some work has focused on the extent to which human vocalizations follow predictable acoustic *forms* that parallel their evolved *functions* (Darwin, 1872; Ohala, 1995; Pisanski *et al.*, 2022). A cry of pain, for example, is

---

predicted to be high-pitched, loud, and acoustically harsh to grab the attention of listeners and elicit aid (Pisanski *et al.*, 2022).

Comparatively few studies have examined whether form-function mappings may also give rise to systematic, predictable vowel patterns in vocalizations. Decades ago, Ohala (1984, 1995) predicted that in animal and human communication, aggressive vocalizations may be accompanied by protruding lips giving rise to what he termed the "o-face," whereas submission may be signaled by drawing the lips back, as in a smile. These articulatory gestures would evoke specific vowel patterns such as more o-like vowels in aggression and more i-like vowels in submission. More recent work mapping the human acoustic space further suggests that vocalizations are poorly articulated relative to speech, meaning that they are produced with limited manipulation of the vocal tract and articulators (lips, jaw, tongue) and thus utilize a narrower vowel space, generally containing a high proportion of open a-like vowels, probably owing to wide mouth opening (Anikin *et al.*, 2023).

In contrast to nonlinguistic vocalizations, which bear no word-like form, interjections are traditionally defined by linguists as the class of words that do not combine with the rest of the grammar and are instead often used as "standalone" units in communication (Wilkins, 1992). Specifically, "expressive" interjections (Ameka, 1992) are the subset of interjections used to communicate a speaker's states, attitudes, or experiences that we will, to be concise, group together here under the label "emotion." Examples of expressive interjections are *ouch* and *ay*, which typically communicate pain in English and Spanish, respectively. Such expressive interjections can therefore be regarded as conventionalized alternatives to screams, cries, grunts, moans, and other nonlinguistic vocalizations.

If some expressive interjections are indeed "linguistically conventionalized vocalizations"—either transformed (perhaps more "controlled") or imitative (arising from imitation of vocalizations expressing the same affect)—then one could predict that their forms, across the world's languages, may bear acoustic resemblances with the nonlinguistic vocalizations used to express the same emotional states. In this spirit, Dingemanse (2023) suggests that "*some interjections can be linked to ancestral vocalizations or bodily responses. Pain interjections provide an instructive example. Most spoken languages appear to make available a pain interjection that has as its nucleus and prosodic peak an open central unrounded vowel. It is hard to escape the conclusion that such forms harken back to a common mammalian pain vocalization*" (Darwin, 1872; Ehlich, 1985)."

To begin to untangle these ideas, in this study, we consider interjections expressing three distinct emotional experiences—pain, disgust, and joy—across up to 131 languages (more than 600 tokens) and compare these with nearly 500 vowel segments from nonlinguistic vocalizations expressing the same emotions, audio recorded from hundreds of speakers of five languages. In line with

Dingemanse's hypothesis, we focus on vowels, which offer a methodologically manageable acoustic space for this initial investigation, and which have been largely overlooked in research on nonlinguistic vocalizations.

With respect to vocalizations, we predict that pain vocalizations should most often be produced with a wide-open mouth, which may give rise to a higher proportion of low vowels such as [a] and [ɐ]. [1] Arguably, joyful vocalizations may more often be produced with a smile, echoing the predictions of Ohala (1995), which would give rise to a higher proportion of mid and high front vowels such as [i] and [e]. In contrast, disgust vocalizations often arise in response to repulsive stimuli such as rotting food or indices of disease and may, even in their volitional conventionalized form, still echo reflexive responses such as gagging that could give rise to a higher proportion of central schwa-like vowels produced with little to no articulation (for more on disgust vocalizations, see also Scherer, 2019, p. 61). While these examples link vocal sounds to reflexive or physical states, sound-symbolic sensory associations, as noted in the following, may also influence the acoustic structures of vocal signals. Most notably, positively valenced vocal sounds, such as those produced in response to joy and pleasure, may contain a higher proportion of "bright" vowels such as [i] and [e] (Butcher, 1974) compared to negatively valenced vocal sounds produced to express pain or disgust.

If expressive interjections are related in form to their nonlinguistic vocal counterparts, the same logic may explain differences in the vowel spaces of interjections expressing pain, disgust, and joy. Furthermore, even if expressive interjections do not recruit the same mechanisms hypothesized here, interjections may simply *imitate* nonlinguistic vocalizations. In any case, in this study, we ask whether features of nonlinguistic vocalizations are mirrored in lexical interjections, which would then be iconic in the broad sense of being non-arbitrary.

We focus exclusively on vowels which we retrieved from interjections transcribed in dictionaries and lexical databases, and from the formant frequencies measured directly from audio recordings of nonverbal vocalizations. Formant frequencies represent vocal tract resonances and, as described in Secs. II A, II B, III A, and III B, the first two formants ($F1$ and $F2$) largely determine the vowel quality of a vocal sound and allow us to characterize vowels in vocalizations (Behrman, 2007). It should be noted that acoustic and prosodic aspects of interjections and vocalizations, such as the fundamental frequency (perceived as voice pitch) and its dynamic variation across an utterance, are also likely to vary across emotional contexts and to carry important functional information, as already highlighted by existing research on nonverbal vocalizations and speech (for reviews see Pisanski *et al.*, 2022 and Scherer, 2021). Here, we take a unique approach by focusing not on prosody and pitch, but rather on vowel quality, and the extent to which vowels may also encode functional information across both interjections and vocalizations.

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.*   3119

B. State of the art

Linguistics has produced a wealth of literature on iconicity or "sound symbolism" in human languages (see, for instance, Hinton *et al.*, 2006; Reilly *et al.*, 2008; De Carolis *et al.*, 2017, among many others). As early as 1929, Sapir (1929) already wrote about "phonetic symbolism," and in the 1960s researchers showed that specific vowel sounds are often mapped to specific perceptual dimensions (Fischer-Jørgensen, 1968), such as in the classic Bouba-Kiki effect, a strong sound-symbolic correspondence that appears to be present even in toddlers (Spector and Maurer, 2013). Some studies have focused on imitative linguistic resources, i.e., ideophones (e.g., Voeltz and Kilian-Hatz, 2001; Dingemanse, 2011, 2023; Reiter, 2012; Haiman, 2018), and others on expressive dimensions (e.g., Bergen, 2004; Vallery and Lemmens, 2021 on slurs). Yet, virtually nothing is known about the role of iconicity in interjections. This may seem surprising given intuitive reasons to expect that expressive interjections may indeed be partly iconic, and because they may offer a case where the potential source of iconicity—influence from nonlinguistic vocalizations—is relatively self-evident, yet also largely untested. The best explanation for this gap is probably the combined scarcity of studies on interjections (Dingemanse, 2017, 2024; Colombat and Lahaussois, 2019) and nonlinguistic vocalizations in adults, though research on vocalizations is mounting rapidly (e.g., Anikin *et al.*, 2018; Anikin *et al.*, 2023; Anikin *et al.*, 2024; Anikin and Lima, 2018; Brooks *et al.*, 2023; Cowen *et al.*, 2019; Ćwiek *et al.*, 2021; Laukka and Elfenbein, 2021; Sauter *et al.*, 2010; Scherer, 2019; Valente *et al.*, 2025; and see Kamiloğlu *et al.*, 2020 and Pisanski *et al.*, 2022 for recent reviews). Notably, extremely few studies directly compare vocalizations and interjections.

How, then, can we assess whether the forms of interjections across the world's languages are indeed iconic, that is, non-arbitrary? If the forms of some interjections were somehow influenced by those of nonlinguistic vocalizations, then formal resemblances should still be detectable in at least some interjections in contemporary languages across the world. Detecting and mapping such resemblances in a comprehensive and systematic manner is a long-term research program. Studies examining the extent to which nonlinguistic vocalizations share similarities across cultures in both production and perception are on the rise, but still scarce. There is emerging evidence that form-function relationships or iconicity in vocalizations may be preserved to some extent across diverse populations. Indeed, the acoustic forms of vocalizations appear to share some emotion-specific acoustic characteristics, and their intended emotions can often be identified by listeners, with some degree of accuracy across cultures (see Ćwiek *et al.*, 2021; Brooks *et al.*, 2023; Sauter *et al.*, 2010; Pisanski *et al.*, 2022, for review). However, studies of vocalizations involving acoustic analyses have mostly focused on acoustic parameters such as fundamental frequency (voice pitch) (e.g., Schwartz and Gouzoules 2019),

loudness (e.g., Anikin *et al.*, 2024), or acoustic harshness arising from nonlinear phenomena (Anikin *et al.*, 2020). Further, studies that have measured formants in vocalizations have focused on absolute formant spacing, an index of vocal tract length and thus body size (Charlton *et al.*, 2020), rather than on vowel qualities emerging from formants (but see Anikin *et al.*, 2023). It thus remains largely unknown the extent to which vowel patterns are preserved in emotional vocalizations or interjections across languages and cultures. Also, to our knowledge, no study has ever considered the resemblance between spontaneous vocalizations and conventionalized interjections.

C. Research questions

Here, following the hypothesis proposed by Dingemanse (2023) noted previously, we focus on vowels, which have high incidence in nonlinguistic vocalizations (see Sec. II C 2). Our research questions are as follows: (1) Do interjections for a given emotional state exhibit resemblances in forms across a large sample of the world's languages? (2) Are certain vowels more prevalent in interjections across languages, and if so, which vowels? In other words, is there a "vowel signature" specific to interjections expressing different emotions, and what does it look like? Finally, (3) Can the vocalic tendencies observed in interjections be ostensibly traced back to nonlinguistic vocalizations?

In this study, we aim to offer a first quantitative insight into this important yet overlooked aspect of human communication and to pave the way for future research on the origins and mechanisms that shape interjections in the world's languages. As an entry point into our research questions, we analyze around 600 interjections in 131 languages and nearly 500 vowels from nonlinguistic vocalizations across five languages, noting that our data are diverse but not equally balanced across languages. An overview of our data and analysis workflow is given in Fig. 1.[2] Research questions 1 and 2 are addressed in a study on interjections in Sec. II, where we give evidence that pain interjections share formal resemblances across languages, including a strong prevalence of a-like vowels and certain diphthongs, whereas global vowel signatures are less apparent in disgust and joy interjections. Research question 3 is addressed in a study on nonlinguistic vocalizations in Sec. III, where our results suggest that all three emotion contexts yield a distinctive vowel signature. Like pain interjections, pain vocalizations have more a-like vowels, whereas disgust vocalizations have more schwa-like central vowels and joy vocalizations have slightly more i-like vowels than expected by chance. Finally, in Sec. IV we compare our results for interjections and nonlinguistic vocalizations, discussing the implications of these results with respect to the iconicity of interjections.

II. INTERJECTIONS

A. Data

We first introduce the methodology implemented to build the interjections dataset. Because our research

3120   J. Acoust. Soc. Am. **156** (5), November 2024
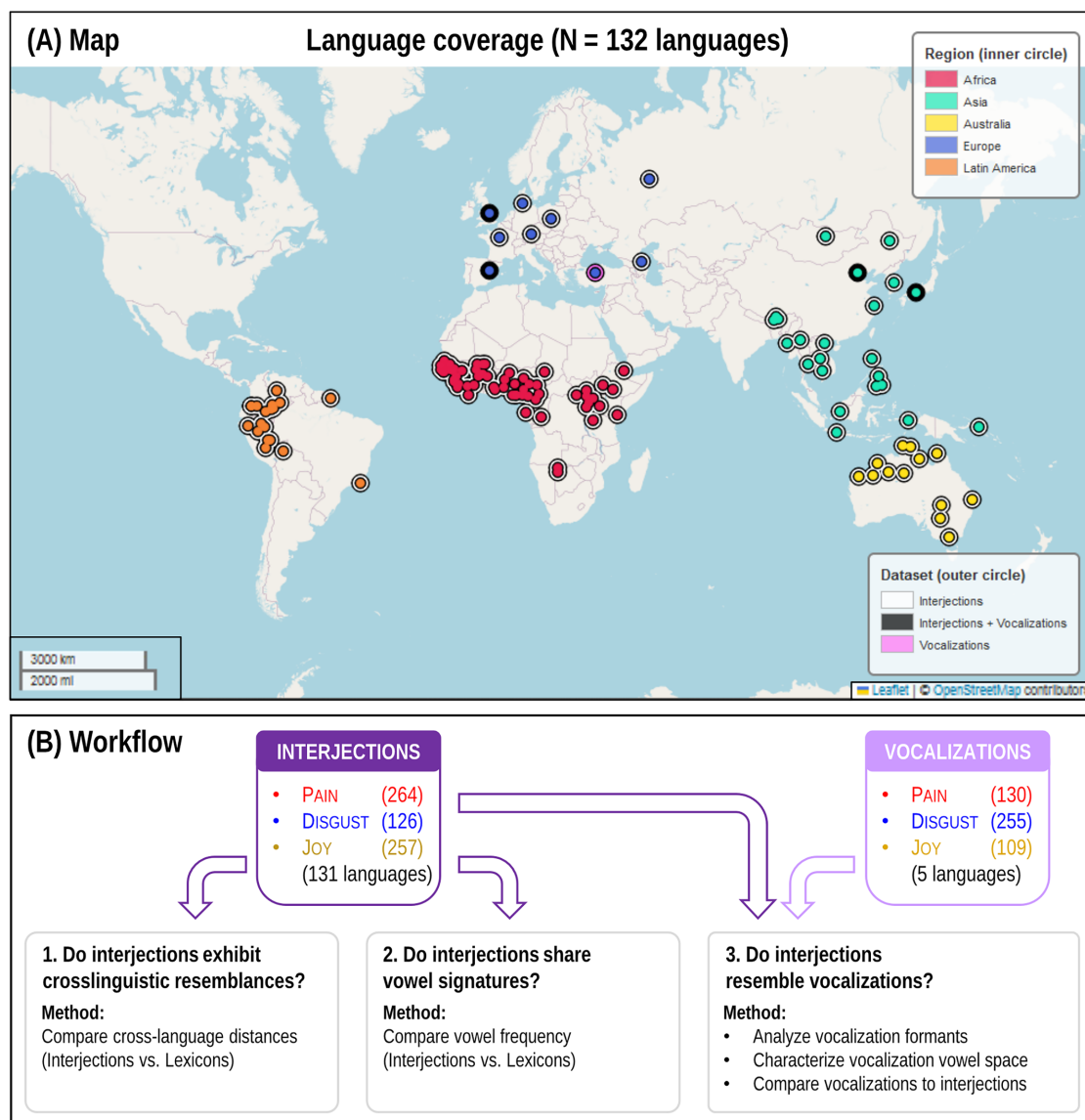
Ponsonnet *et al.*

FIG. 1. (Color online) General overview of data and analyses. (A) Locations of the languages represented in the interjection and/or vocalization datasets (map drawn with the lingtypology Python package, Voronov and Forkel, 2019). See SuppPub1.HTML for an interactive version (B) Outline of the three key research questions and datasets.

questions involve the detection of potential specificities of interjections, we subsequently introduce the lexical datasets that provided a baseline against which this detection was performed. Finally, we provide information on the transcription framework adopted and the general method implemented to compare interjections to their lexical counterparts.

1. Interjection dataset

Our dataset of lexical interjections consists of 647 interjections that express pain, disgust, or joy, from up to 131 distinct languages [Fig. 1(A)], gathered from dictionaries and lexical databases. This dataset is made available in the supplementary material as file SuppPub2.csv (please note that this file is encoded in utf8 when opening it).

a. Emotional categories: Pain, disgust, and joy. Pain was chosen because pain interjections are often treated as prototypical—as in Dingemanse's hypothesis noted previously (Dingemanse, 2023). In addition, a typological study of Australian interjections recently revealed that pain interjections can be shared across large numbers of geographically and phylogenetically distant languages (Ponsonnet, 2023). Disgust and joy were chosen as counterparts or controls to pain for each emotional valence, negative and positive, to make sure that the features we observed for pain interjections were not either properties of expressive interjections in general or properties of interjections expressing negative experiences. Other categories such as fear or surprise were considered but deemed unsuitable because pilot investigations indicated that they more rarely correspond to well-identified interjections in lexicographic sources

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.* 3121

(i.e., dictionaries). For instance, interjections that express surprise can typically express pain, joy, or fear at the same time, and this would have blurred our observations. Note that there is some overlap amongst interjections for pain, disgust, and joy as well, in the sense that some interjections express two (or even all) of these emotions together. However, these overlaps represent less than 5% of all interjection tokens.

Without presuming that members of all cultures, or speakers of all languages across the world, embrace the same emotion categories (see, for instance, Harré, 1986; Wierzbicka, 1999; Feldman Barrett, 2009), we defined operational categories for the purpose of lexical data collection. Pain was defined as physical pain, and interjections expressing solely emotional suffering were excluded (e.g., in English we included *ouch* but excluded *alas*). Disgust was defined as disgust of physical stimuli as opposed to social disgust or moral contempt (e.g., in English we included *yuck*, but excluded *pfff* which can be a vocal marker of social disapproval). Interjections qualified for joy if there was sufficient evidence that they expressed an intense satisfaction and, thus were both intense and positive. This excluded interjections for mild satisfaction (e.g., *good*, *okay*) and interjections for surprise, which may be intense but neutral rather than positive. The joy category thus encompasses rejoicing (e.g., *yay*), interjections of congratulations (e.g., *bravo*), etc., as well as interjections such as *wow* in cases where the source suggested intensity (e.g., use of exclamative punctuation). The joy category is more heterogeneous than pain and disgust are, and this mirrors its lexicographic heterogeneity: while a significant proportion of our sources featured interjections described simply as "pain interjection" or "disgust interjection" (or equivalent glosses), the descriptions of the interjections we categorized as joy interjections were more diverse. We managed this thanks to a number of systematic rules concerning whether to include certain tokens or not (e.g., interjections expressing simple amazement, or congratulations without satisfaction, were excluded). This heterogeneity does not hinder the validity of the joy interjections set as a control for the patterns observed in pain interjections.

b. Language sample. Because interjections have long remained under-studied and under-documented in linguistics (Dingemanse, 2017, 2023; Colombat and Lahaussois, 2019), our language sample was partly dictated by limitations in accessing the data. In other words, we were unable to constitute what would be regarded as a "balanced" sample in linguistic typology, that is, a sample whose members replicate the geographic and genealogical distributions of languages across the world. Given the current empirical data available on interjections, any typological study of interjections can only be carried out within the framework of this limitation. With this in mind, throughout this study, our analyses focus on robust, concordant effects.

While our sample is not fully balanced, it is very diverse. To ensure such geographic and phylogenetic diversity in the data, we investigated sources and collected data from 131 languages across five different regions of the world: Africa, Asia, Australia, Europe, and Latin America. Again, the data are not evenly distributed across these regions, families, languages, and emotions (e.g., no disgust interjections were found in Australian languages, in line with the low degree of lexicalization of this emotion on the continent, see Ponsonnet and Laginha, 2020, p. 28). Descriptive statistics are available in the supplementary material (Table S1 in SuppPub3.pdf). Within each region, we avoided pairs that were genetically close enough to be regarded as variants of the same language and made sure to include a range of distinct families and subgroups. Based on the Glottolog classification (Hammarström *et al.*, 2024), our sample includes 45 families and 17 isolates or languages from unknown groups. Therefore, our sample is sufficiently diverse to avoid biases towards a specific part of the world or a specific profile of languages.

For most regions, we used digital versions of dictionaries, either as PDFs or collated online (particularly for Asian languages, for which we accessed these sources via www.sealang.net). For African languages, we were able to harvest data from the RefLex database (Segerer and Flavier, 2023). All references are provided in the supplementary material (SuppPub4.ods).

c. Methodology of interjection data collection. Our sources for interjections were principally published dictionaries, using mostly English as a metalanguage, although most dictionaries for Latin American languages were in Spanish and a few in Portuguese. We also worked with French and German as metalanguages when investigating European dictionaries. For each source, we systematically carried out a series of automatic text searches. We first looked for interjections[3] generally and collected all the hits matching the definitions for our three emotional categories. We then searched for the interjections in the metalanguage of the source (e.g., *ouch* for dictionaries in English, *ay* for dictionaries in Spanish) and for a series of keywords that would take us to sections of the source featuring items of interest (e.g., for disgust: "repugnance," "repulsion," "gross," "dirty," etc.). This tedious but necessary process allowed us to identify relevant tokens that had not been flagged as interjections by the authors of the dictionaries (which was not uncommon, given that dictionary makers are not necessarily attentive to the exact nature of these lexical items, see footnote 3).

We collected all forms listed as interjections, including those with another meaning in the same language, such as *wónā* in the Mundu language (North Volta-Congo, Niger-Congo), a pain interjection which is also the word for "mother."[4] For each hit, all the interjection variants listed in the source were included in our dataset, even if they were similar in form, such as *akatai* and *akatsai* for pain in the Cocama language (Tupian, Latin America). This remained the most neutral solution given the lack of criteria to decide when two pronunciations should be regarded as variants of a unique interjection, or as different interjections.

16 November 2024 15:11:14

TABLE I. Examples of interjections with their transcription in the simplified ASJPcode.

Region	Glottocode	Language	Emotion	IPA transcription	ASJP Transcription
Africa	sere1260	Sereer	Joy	suuʔ^{j}in n	suu7inn
Africa	cent2050	Kanuri	Pain	wájjájóòʔ	wayyayoo7
Asia	kore1280	Korean	Joy	ejla t͡ɕoh kʰwun a	eylaCohkwuna
Europe	stan1293	English	Disgust	jɪəx	yi3x
Europe	stan1288	Spanish	Joy	t͡ʃuta	Cuta

2. Transcriptions: IPA and ASJPcode

Where not available directly from the entries in the sources, transcriptions in the International Phonetic Alphabet (IPA) were deduced based on the orthographic information in the source, crossed with the phonemic inventory and alphabetic language for each language/ source, and verification with language experts when needed. In this process, we deliberately omitted prosodic features (e.g., tone, vowel lengthening). Due to the overwhelming prosodic versatility of interjections in usage, these are far less stable than segmental features, and likely less conventionalized.

The IPA is the most accurate and standardized transcription system, but it is sometimes too fine-grained for statistical detection of cross-linguistic similarities. We eventually aim to assess the role of targeted articulatory features. In this study, however, as a first step, we chose to follow a methodology already applied in the context of historical linguistics or for cross-linguistic comparisons of sound-meaning associations (e.g., Blasi *et al.*, 2016; Erben Johansson *et al.*, 2020). We thus used the ASJP framework, which provides a simplified transcription system named ASJPcode, developed for the Automated Similarity Judgment Program (Brown *et al.*, 2008; Brown *et al.*, 2013). Each ASJPcode symbol can represent one or several IPA vocalic or consonantal phonemes. More specifically, the ASJP coding divides the vowel space into seven categories by lumping together several vowel qualities. For example, the ASJP character "a" stands for both [a] and [ɐ], while "i" corresponds to [i, y, ɪ, ʏ] (see Table S2 in SuppPub3.pdf). This coarse-grained transcription is particularly suitable for computing lexical distances in the context of this study because our hypotheses investigate potential differences between interjections and lexicons, rather than their fine-grained phonetic composition.

The interjections collected and their lexical counterparts were thus converted from IPA into ASJPcode; secondary articulations as well as tones were removed with a tailored-made PYTHON script. A few examples of this process are reported in Table I. In the main text of the paper, all IPA vowels are indicated with square brackets (e.g., [a]), whereas ASPJ vowels are indicated with quotes (e.g., "a").

3. Lexical databases

Our first aim was to determine whether interjections are characterized by a recognizable "vowel signature" when compared to the general lexicon in each language. Here, the notion of vowel signature refers to potential patterns of higher or lower proportions of specific vowels, which may be more or less common in interjections compared to the other words of a given language. To test this, we retrieved lexical data from two online resources, the ASJP database (Wichmann *et al.*, 2022) and the Lexibank database (List *et al.*, 2022). These lexical baselines allowed for an additional control, wherein we can compare the proportions of vowels not only across interjections for the three emotions, but also compared to the proportions of different vowels found in the languages at large.

The ASJP and Lexibank databases have their own pros and cons, hence the benefit of using both. The ASJP database provides translations of 40 basic concepts drawn from the Swadesh list (e.g., one, two, eye, sun) in more than 5500 languages, and is often used to compute lexical distances across languages; whereas Lexibank is a unified framework aggregating lexical datasets for about 2000 languages. The lexicons made available in Lexibank cover about 300 meanings on average, thus providing a more precise snapshot of the lexicons of the world's languages at the expense of a narrower language coverage.

Despite their impressive lists of languages, both ASJP and Lexibank present a partial overlap with the interjection dataset we collected. Out of the 131 languages in the interjection dataset, 126 languages are also present in ASJP and 69 languages appear in Lexibank. The statistical analyses of interjections (see Sec. II B) were replicated with both lexical datasets, adopting the same ASJP transcription scheme as for the interjection dataset. Convergent results obtained with both the basic ASJP lexicon and the more expansive Lexibank lexicon were considered robust. ASJP wordlists were retrieved from https://github.com/lexibank/asjp (version 20, Wichmann *et al.*, 2022) and Lexibank lexicons from https://github.com/lexibank/lexibank-analysed (version 1.0, List *et al.*, 2022).

Some of the languages for which we collected interjections are absent from the lexical databases: five of our initial 131 languages with interjections are not found in ASJP, and 61 are not found in Lexibank. Once these languages were discarded, we were left with 636 interjections and 17 211 other words when considering ASJP; and 413 interjections and 36 243 other words when considering Lexibank.

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.* 3123

4. Associating interjections with length-matched words from the lexicon

Interjections and other lexical items vary in their number of phonological units (their "length"), and contrasting them without controlling for this factor may lead to biases in the results. We therefore matched each interjection with lexical entries from the same language featuring the same number of phonological units. For each language, this approach resulted in ignoring (i) interjections without length-matched lexical entries, and (ii) lexical entries whose length did not match any interjection. In the case of the ASJP dataset, 588 interjections and 6987 other words remain, while in the case of Lexibank, 393 interjections and 15 318 lexical entries remain.

This procedure gave us solid ground to assess the specific properties of interjections, while controlling for their length. More details about the matching process and the discarded entries are provided in the Supplementary Material (see Sec. 5 in SuppPub5.HTML and SuppPub6.HTML).

Data preparation and conversion to simplified ASJP were implemented in Python, using the asjp library (version 0.02 https://github.com/pavelsof/asjp). Matching between interjections and lexical entries was implemented in R. The PYTHON and R/markdown codes are provided in the SuppPub7.zip archive.

B. Analyses

1. Imbalances in the dataset, subsampling approach, and statistical inference

To assess whether interjections tend to show distinct vowel signatures, we implemented a procedure based on a repeated subsampling of our data. It takes advantage of the relatively large amount of data available while controlling for the unavoidable imbalance in (i) the language families represented in the dataset compared to the distribution of the world's languages and (ii) the numbers of interjections collected in each language.

According to Glottolog (Hammarström *et al.*, 2024), the languages in the dataset belong to 45 families plus 17 isolates or languages from unknown groups, giving $45 + 17 = 72$ stocks (details are given in Sec. 6.1.4 of the SuppPub5.HTML and SuppPub6.HTML supplementary material). We have for instance 19 North-Central Atlantic languages (a branch of the larger Atlantic-Congo family), each represented in our dataset by between 0 and 6 interjections of pain, 0 and 3 interjections of disgust, and 0 and 8 interjections of joy.

For each emotion, we built 1000 subsamples as follows: for each subsample, we randomly selected in each stock one language with at least one interjection for the target emotion, and for this language, we selected one of these interjections, and one length-matched word from the lexicon. This procedure gives in each subsample an equal weight to each language stock (families or isolates) whatever the number of languages in the family and the number of interjections in the language.

As languages belonging to the small families represented in the dataset contribute more to the global analysis because they are drawn more often (or even every time for isolates and families represented by a unique language), we also implemented an additional mitigation procedure: one-quarter of the languages were randomly removed from each subsample, which reduced the previously noted incidental over-representation. In a few samples, dropping languages randomly resulted in the absence of the two regions with the least languages: Europe and Australia. This did not impact the statistical approach over the 1000 subsamples in any significant manner.

Given the 72 language stocks from which the samples are drawn, this procedure provides a robust way to statistically detect distinct interjection signatures while controlling for the genealogical structure of the dataset. To reduce the risk of artificially inflating the number of signatures detected, we adopted a conservative approach by considering signatures that could be simultaneously detected in multiple regions, i.e., in the absence of language contact. We did not apply a stricter control for areal contact within each region because of the limited size of the dataset.

To determine whether the interjections tend to differ from their matched lexicons in terms of cross-linguistic distance (research question 1) and vowel content (research question 2), we relied on a non-parametric permutation approach for comparing the interjection and lexicon distributions across the subsamples.[5] To perform statistical inference and obtain *p*-values, a *t* value derived from two paired distributions for interjections and other words (derived from 1000 subsamples) was compared to a distribution of 1000 *t* values each obtained from 1000 samples where the two elements in each pair of interjection and length-matched word were randomly permuted. These permuted samples constituted a baseline in which interjections are statistically indistinguishable from the lexicon because of the random permutation. They were drawn following the same rules as previously described (one language per language family/ stock, one interjection/lexicon pair per language, pruning one-quarter of the selected entries), before the additional random shuffling of interjections and lexical entries. The empirical *p*-value was equal to the number of permuted samples whose *t*-value was more extreme than that of the target subsample, with a null hypothesis of no difference between interjections and their lexical counterparts.

The strength of the various effects (the "effect size") we investigated was quantified with Cohen's d (Cohen, 1969; see further descriptions). All statistical analyses were implemented in R. The full procedure is detailed in the supplementary material (see Sec. 7 in SuppPub5.HTML for ASJP, and SuppPub6.HTML for Lexibank) and the corresponding R/ markdown codes are provided in the SuppPub7.zip archive.

2. Measuring cross-language resemblances in interjections

We tested whether, for each emotion, the interjections observed were closer (convergence, indicating similarity in

3124   J. Acoust. Soc. Am. **156** (5), November 2024
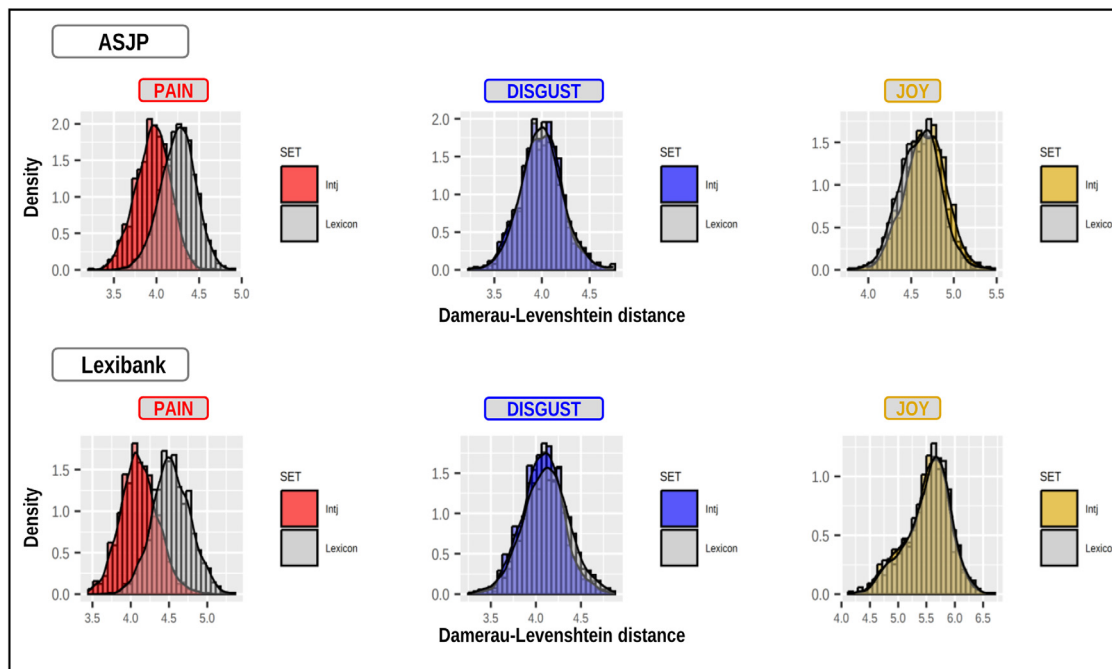
Ponsonnet *et al.*

FIG. 2. (Color online) The cross-linguistic distance between interjections varies by emotion. For each emotion, distributions of the Damerau-Levenshtein distances were computed with the interjections (colored distributions) and lexicons (gray distributions) sampled from the ASJP dataset (top) and Lexibank dataset (bottom). The distributions show that pain interjections are highly similar to one another across languages, much more so than to other words in the same language. Note that scales differ across the panels.

forms), farther away (divergence, indicating differences in forms), or as distant as the other words in the lexicons across the languages under study. More precisely, for each emotion in each of 1000 subsamples, we computed the average cross-linguistic Damerau-Levenshtein distances (Damerau, 1964; Levenshtein, 1966) between all the selected interjections. The procedure was then repeated with the length-matched lexical items. The Damerau-Levenshtein distance is defined as the minimum cost of operations (insertions, deletions, substitutions, or transpositions) required to transform a phonological sequence A into a sequence B (here we assume that all these operations share the same cost of 1).[6] We considered all segments to compute operations and associated costs, i.e., both vowels and consonants. For each emotion, the distributions of the cross-linguistic distances for the interjections and their lexical counterparts over the subsamples were statistically compared and an empirical p-value based on permutation tests was computed, as well as an effect size.[7]

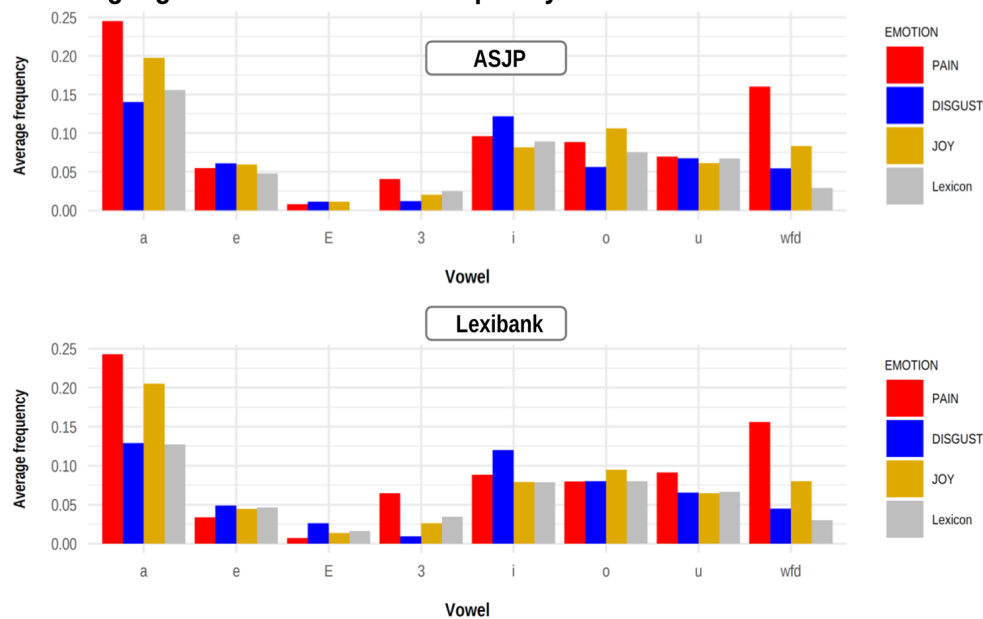### 3. Determining the vowel signatures of interjections

In this analysis, we focus on the composition of interjections in terms of ASJP vowel categories. The "E" vowel category (low front vowels, rounded and unrounded) was deliberately omitted because of its low overall frequency in both interjections and lexicons (see SuppPub5.HTML and SuppPub6.HTML, Sec. 9 for details), leaving us with six monophthong vowel categories from the ASJP: "a," "e," "3" (high and mid central vowels), "i," "o," and "u."

Beyond the monophthongal vowel composition of interjections, we also investigated a partly anecdotal, yet impressionistically notable and intriguing observation: several languages used expressive interjections that involved a diphthong ("gliding vowels") or a vowel-semivowel sequence, and more specifically a wide falling diphthong, defined as a trajectory starting with a low vowel and ending with a high vowel or semivowel and characterized by a decreasing prominence (Jones, 1954). For instance, pain interjections in English (*ouch*, [aʊt͡ʃ]), French (*aïe*, [aj]), Japanese (痛い, [itai]), Mandarin Chinese (哎哟, [aːi jo]) and several Australian languages (e.g., Marrthunira *yakayi*, [jakaji]) are all characterized by a high proportion of such wide falling diphthongs, as also underlined by Ponsonnet (2023). In terms of ASJPcode, these diphthongs correspond to one of these sequences: "ai," "a3," "au," "ay," and "aw" (or in

TABLE II. Comparison of cross-linguistic distances between interjections and lexicons. The statistical analysis rejects the null hypotheses that distances follow the same distributions for interjections and lexicons. However, only pain interjections consistently show more resemblance than their lexical counterparts for both datasets, with a very large effect size. For disgust and joy interjections, the effect size is mostly very small and may be interpreted as an artifact due to the analysis sensitivity rather than a meaningful effect.

Emotion	ASJP dataset (123 languages)		Lexibank dataset (69 languages)	
	p-value	Effect size	p-value	Effect size
Pain	< 0.001	1.47 (very large)	<0.001	1.71 (very large)
Disgust	< 0.001	0.04 (very small)	<0.001	0.16 (very small)
Joy	< 0.001	0.25 (small)	<0.001	0.06 (very small)

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.*   3125

(A) Cross-language overview of vowel frequency

ASJP

Lexibank

Average frequency vs *Vowel* (a, e, E, 3, i, o, u, wfd)

EMOTION: PAIN, DISGUST, JOY, Lexicon

(B) Vowel signatures

B1) ASJP DATASET (126 LANGUAGES)

	INTERJECTIONS: EMOTION													
	DISGUST				JOY					PAIN				
VOWEL	Africa	Asia	Europe	Latin America	Africa	Asia	Europe	Latin America	Australia	Africa	Asia	Europe	Latin America	Australia
a	++	--	-			++		--	--	++	++	+	++	++
e	--	+	-		+		++	-		++			-	
3	--		-	-	-			++		--	+		++	
i	+	++		+	--		-		++	-	+			
o			-		++			++	++					-
u	--	--	++				++	--	--	-		++	--	--
Diphthong	++			++	++	++		++	++	++	++	++	++	++

B2) LEXIBANK DATASET (70 LANGUAGES)

	INTERJECTIONS: EMOTION													
	DISGUST				JOY					PAIN				
VOWEL	Africa	Asia	Europe	Latin America	Africa	Asia	Europe	Latin America	Australia	Africa	Asia	Europe	Latin America	Australia
a	++					++				++	++	++	++	++
e	--		-	+	+				-					+
3			-	--			++				+		+	
i	++	++		++			-		++	--	++			
o	++	--	-	--	+				++	+	-	-		-
u	--		++				++	--	--			++	--	--
Diphthong	++	++		++		++	+	+	++	++	++	++	++	++

B3) CONSENSUS SUMMARY

	INTERJECTIONS: EMOTION													
	DISGUST				JOY					PAIN				
VOWEL	Africa	Asia	Europe	Latin America	Africa	Asia	Europe	Latin America	Australia	Africa	Asia	Europe	Latin America	Australia
a	++					++				++	++	+	++	++
e	--		-		+				-					
3			-	-				++				+		+
i	+	++		+			-		++	-	+			
o	++	--			+				++					-
u	--		++				++	--	--			++	--	--
Diphthong	++			++	++			+	++	++	++	++	++	++

FIG. 3. (Color online) Vowel signatures in emotive interjections across regions of the world (A) Vowel frequencies by emotion and in the lexicon, where "wfd" stands for wide falling diphthong. (B) Vowel signatures for each Emotion Category and Region (by column) and each vowel (by row). The three sub-panels B1, B2, and B3 are organized in the same way: each cell corresponds to the comparison of frequency distribution (interjections versus lexicons) for a given vowel (organized by row), region, and emotion (organized by columns). (B1): ASJP dataset. (B2) Lexibank dataset. (B3) Consensus summary of robust results present in both (B1) and (B2). Symbols: $++$ indicates much more frequent in interjections than in lexicons (p-values <0.05 and large effect size). $+$ indicates more frequent in interjections than in lexicons (p-values <0.05 and medium effect size). $--$ indicates much less frequent in interjections than in lexicons (p-values <0.05 and large effect size). $-$ indicates less frequent in interjections than in lexicons (p-values <0.05 and medium effect size). Empty cells correspond to no statistical difference between interjections and lexicons or to small effect sizes. Note that no disgust interjections were found in the Australian languages we surveyed and that the vowel category "E" is absent due to a low overall frequency in the language samples, as illustrated in (A). All vowels are described in terms of ASJPcode, however, quotes are omitted for legibility purposes.

IPA: [ai], [aj], [aə], [aɜ], [au], [aɯ], [ɐi], etc.). To test whether this pattern of falling diphthongs is widespread and could indeed be a signature of pain interjections (and potentially other categories of expressive interjections), we also investigated the observed frequency of diphthongs compared to the baseline lexicons. Unless otherwise stated, in the remainder of this article the term "vowel" in the context of interjections includes both the monophthongs and these wide falling diphthongs.

To investigate whether interjections are characterized by specific vowel signatures, we assessed whether the hypothesis that they do not differ from the statistical distribution observed in their length-matched lexicons (null hypothesis) can be rejected. For each emotion, we based our analysis on the same 1000 subsamples described in the previous section. The vowel frequency[8] observed in the interjections was then compared to the lexical distribution and an empirical p-value was computed, as well as an effect size (again, with the same approach as for the distances). We consider that a vowel signature is attested when the observed frequency in interjections is significantly larger (preference) or smaller (avoidance) than the lexical baseline with an associated effect size interpreted as at least medium (Cohen's d $\geq$ 0.5). On the contrary, when the null hypothesis cannot be rejected, or when the size effect is very small or small, we conclude an absence of a specific vowel signature for the given set of interjections. This conservative procedure, based on the effect size rather than the simple p-value, is consistent with the exploratory nature of our study and the focus on robust effects. For similar reasons, we discuss only vowel signatures that are present in both the ASJP and Lexibank datasets, to further avoid overinterpreting potentially spurious or small effects.

The previous computations and statistical tests were based on a frequency of occurrence of vowels defined as the ratio of the number of occurrences of each vowel in an interjection or a lexical word over the length of this interjection or word. To further understand the patterns of prevalence of some vowels in the interjections for pain, disgust or joy, we attempted to decompose them by computing (i) the vocalic frequency of vowels, i.e., the number of occurrences of a vowel in an interjection or a lexical word over the total number of vowels in this interjection or word, and (ii) again in an interjection or word, the ratio of the number of vowels over the total number of phonemes (see SuppPub5.HTML and SuppPub6.HTML, Sec. 9). We did not conduct additional statistical tests for these additional variables.

C. Results and discussion

1. Do interjections exhibit cross-linguistic resemblances?

Results are reported in Fig. 2 and Table II, with both ASJP and Lexibank datasets, which offer two different snapshots in terms of language coverage and lexicon size. In general, both analyses showed similar results: the estimated

effects are of a similar size whether using the ASJP or the Lexibank datasets.

Our analysis is sufficiently sensitive to detect very small differences between the distance distributions computed with the interjections and their matched lexicons (all p-values <0.001). However, a visual inspection shows that despite differences in their shapes, several distributions manifest very similar central tendencies (Fig. 2). For example, the disgust interjections are more similar than their lexical counterparts on ASJP by a value of 0.01 on average (i.e., one-hundredth of a phoneme). This translates into a Cohen's d of only 0.04, which is considered a very small statistical effect. In other words, this tiny absolute difference is probably ecologically meaningless.

In contrast to these small differences, pain interjections exhibit a very large, robust and meaningful cross-linguistic convergence, with much more similar vowel properties than their baseline lexicons (ASJP: d = 1.47 very large effect; Lexibank: d = 1.71 very large effect; Table II). In other words, in our dataset, pain interjections are more similar to one another across languages than they are to the broader lexicon (i.e., non-interjection words). While the average cross-linguistic distance hovers around 4.25 for lexicons, it is close to 3.95 for pain interjections with the ASJP dataset (an average difference of 0.3 phoneme). A similar (in fact slightly larger) difference is observed with Lexibank (4.53 versus 4.11, respectively, see Fig. 2, corresponding to a difference of about 0.4 phoneme).

As an interim conclusion, we show that some expressive interjections exhibit stronger cross-linguistic resemblances than would be expected based on their lexical counterparts. More specifically, a compelling and robust convergence pattern, or robust similarity, is observed for pain. Disgust interjections on the other hand are not effectively more similar to one another than are their lexical counterparts on average, and this is also true for joy interjections, despite small incidental differences in their distributions. Below, we further explore how pain interjections converge by comparing the vowel content of the interjections versus lexicons, testing which vowels appear more (or less) often in interjections expressing pain, disgust, and joy versus in other words found in the languages.

2. Do interjections exhibit distinctive vowel signatures?

Figure 3 summarizes the results of our analyses comparing proportions of different vowels expressing each emotion in interjections and other lexical items across different regions of the world. Figure 3(A) provides an overview of the average frequencies of vowels for each emotion in the interjection dataset, compared to the frequencies observed in the lexicons. Vowel signatures tested statistically by comparing interjections with lexical baselines estimated with the ASJP and Lexibank datasets are presented in Fig. 3(B). The "take home" message is given in panel B3, which summarizes the most robust results that show a consensus between both the ASJP and Lexibank datasets.

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.*   3127

Figure 3(A) illustrates that the baseline frequency varies across vowels, with a general, relative predominance of "a" vowels found in the lexicons and across all interjections, as also previously observed across human vocalizations (Anikin *et al.*, 2023). Similar patterns are obtained with the ASJP and Lexibank datasets. At the same time, the figure also shows potential emotion-specific vowel signatures, with a notably large proportion of "a" and diphthongs in pain interjections as well as, less notably, joy interjections; and a relatively high proportion of "i" in disgust interjections. Further exploration suggests that the large proportion of "a" in pain and joy interjections corresponds to the combination of two factors: a higher prevalence of this vowel compared to other vowels in these interjections, and a higher ratio of vowels in these interjections compared to other words. For "i" in disgust interjections, the situation differs because while this vowel occurs more than other vowels (but not "a"), vowels overall are not more frequent in interjections of disgust than in other words (see SuppPub5.HTML and SuppPub6.HTML, Sec. 9).

Figure 3(B) summarizes the results of the analyses we ran to statistically test for vowel signatures. Empty cells in Fig. 3(B) indicate that the null hypothesis cannot be rejected (the vowel frequency in interjections does not significantly differ from the vowel frequency in lexicons) or differences for which the effect size is very small or small (Cohen's $d < 0.5$). Cells containing a single minus or plus symbol denote a vowel frequency that is significantly lower/larger in interjections compared to what is observed in the lexicons, with a medium effect size ($0.5 \leq$ Cohen's $d < 0.8$). Cells containing double symbols ($++$ or $--$) denote larger effect sizes (Cohen's $d \geq 0.8$). The *p*-values and Cohen's d values for this figure are given in supplementary material (see Sec. 11.4 in SuppPub5.HTML for the ASJP analysis and in SuppPub6.HTML for the Lexibank analysis, respectively) and the frequency distributions are displayed as Figs. S1 and S2 in SuppPub3.pdf.

Focusing on the consensus table (Fig. 3, panel B3), which summarizes the most robust results, we observe that 46 vowel signatures are detected (out of 98 potential signatures—one per cell in the table) and that they mostly involve preference rather than avoidance. In other words, about 70% of vowel signatures (32/46) show that a specific vowel is *more* frequent in interjections than expected based on how often that vowel appears in the language at large, and in 30% of cases, the vowel is less frequent in interjections than in the broader language. For pain and joy interjections, vowels are slightly more frequent in interjections than in the rest of the lexicon (see SuppPub 5, Sec. 9.1).

We also find significant differences across emotions. Pain interjections have the highest number of signatures overall, mostly preference signatures. Two of them, the preference for "a" and for the wide falling diphthong, extend over the five regions represented in our data. This suggests that "a" vowels, and wide falling diphthongs including "a," may be universally more prevalent in pain interjections than in other words. The preference signature of the wide falling

diphthong is also attested in disgust and joy, albeit only for two and three of the regions respectively. In a sense, the vowel signature of joy interjections looks like an "attenuated" version of the vowel signature for pain: eight of the 19 signatures visible in pain are also found in joy. Disgust interjections further differ from pain interjections, and it is notable that compared to pain and joy, disgust exhibits more avoidance signatures, i.e., vowels are relatively less prevalent in disgust interjections than in the rest of the lexicon. However, none of these avoidance signatures are systematic across regions. These results corroborate our Damerau-Levenshtein distance analyses reported in the previous section: pain interjections strongly converge, whereas disgust and joy interjections do not show robust or clear patterns across regions.

Our analyses indicate that vowel distributions also vary across regions of the world, suggesting that the vowel forms of interjections may be influenced by regional factors pertaining to linguistic history and contact (see Ponsonnet, 2023), or in some cases be as random as with non-interjection words. Each region of the world exhibits at least two signatures for each emotion, and, in two cases, almost the whole vowel system exhibits a remarkable pattern, with the language sample from Africa exhibiting six signatures for disgust and the languages from Australia showing five signatures for joy. These regional patterns call for further research.

The existence of regional variation makes the convergence observed in pain interjections even more notable. To summarize, the results show that in our data, pain interjections have a distinctive vowel signature that recurs across the majority of families and languages tested here. They feature more "a" vowels (open vowels) than does the length-matched baseline lexicon across the five world regions considered here. Their "a" vowels more often form wide falling diphthongs (in ASJP "ai," "a3," "au," "ay," and "aw" or in IPA [ai], [aj], [aə], [aɜ], [au], [aɯ]) than observed in the length-matched baseline lexicon across all world regions. The diphthong pattern is even clearer than the "a" vowel pattern, as all effect sizes are very large (Cohen's ≥ 1.2).

Based on our results, therefore, the over-representation of "a" vowels and wide falling diphthongs may be a recurring feature of pain interjections across the world's languages. Of course, this does not mean that any pain interjection in any language will necessarily contain one of these segments. Instead, this indicates that overall, many pain interjections do have these vowel segments, irrespective of their geographic location and genealogical affiliation. The prevalence of "a" comes relatively close to validating the Dingemanse (2023) hypothesis, mentioned in our introduction, that "*[m]ost spoken languages appear to make available a pain interjection that has as its nucleus and prosodic peak an open central unrounded vowel.*" The second part of Dingemanse's suggestion, namely, that "*such forms harken back to a common mammalian pain vocalization,*" we investigate in our analysis of vowel patterns in

nonlinguistic vocalizations in the following, and we come back to this hypothesis in Sec. IV. We may note, however, that Dingemanse's proposal does not mention wide falling diphthongs, although our data and results indicate that in pain interjections, such diphthongs seem particularly salient. We expand on the potential implications of this observation in Sec. IV.

III. NONLINGUISTIC VOCALIZATIONS

A. Data

Our original dataset of nearly 600 vowel segments was derived from 375 volitional nonlinguistic vocalizations intended to express pain, disgust, or joy. These vocalizations were audio recorded and analyzed from a total of 166 men and women with a broad age range, native speakers of one of five languages [Fig. 1(A); SuppPub8.ods]. The languages included Chinese Mandarin (22 men, 23 women, 1 non-binary, aged 20–50), English (24 men, 18 women, aged 18–77), Japanese (9 men, 8 women, aged 21–25), Spanish (15 men, 7 women, aged 20–45), and Turkish (18 men, 21 women, aged 20–51; see Table S5 in SuppPub3.pdf for additional participant details).

1. Vocalization dataset

Participants were recruited via the crowdsourcing platform *Prolific* (2023), except for Japanese participants who were recruited at the University of Tokyo or Ritsumeikan University in Kyoto, and who completed the task in the lab. To take part in the experiments, participants had to report being native speakers of the given language. For Mandarin, Japanese, Spanish and Turkish speakers, some familiarity with English was permitted.

These participants were taking part in a large-scale, cross-cultural experiment in which they were presented with vignettes representing various biologically and/or socially relevant emotional contexts and were instructed to produce a nonlinguistic vocalization, without words, in each context. The emotional contexts used in this paper were chosen to parallel the interjections: pain, disgust, and joy. Specifically, participants were asked to imagine the following scenarios: "*You burn your hand, produce a vocalization to express your pain*" (pain); "*You have eaten some rotten food, produce a vocalization to express your disgust*" (disgust); "*You have won something and you want to celebrate with your family/friends, produce a vocalization to express your achievement*" (joy). All participants performed the task using a custom interface designed in Labvanced (Finger *et al.*, 2016) and were reimbursed monetarily using the recommended rate. Instructions and vignettes originally written in English were translated into the participants' native languages by a native speaker of each language (Japanese, Mandarin Chinese with simplified characters, Spanish, and Turkish), using the back-translation method to ensure cross-linguistic parity.

To ensure that our sample of nonlinguistic vocalizations did not contain any interjections, two researchers coded all
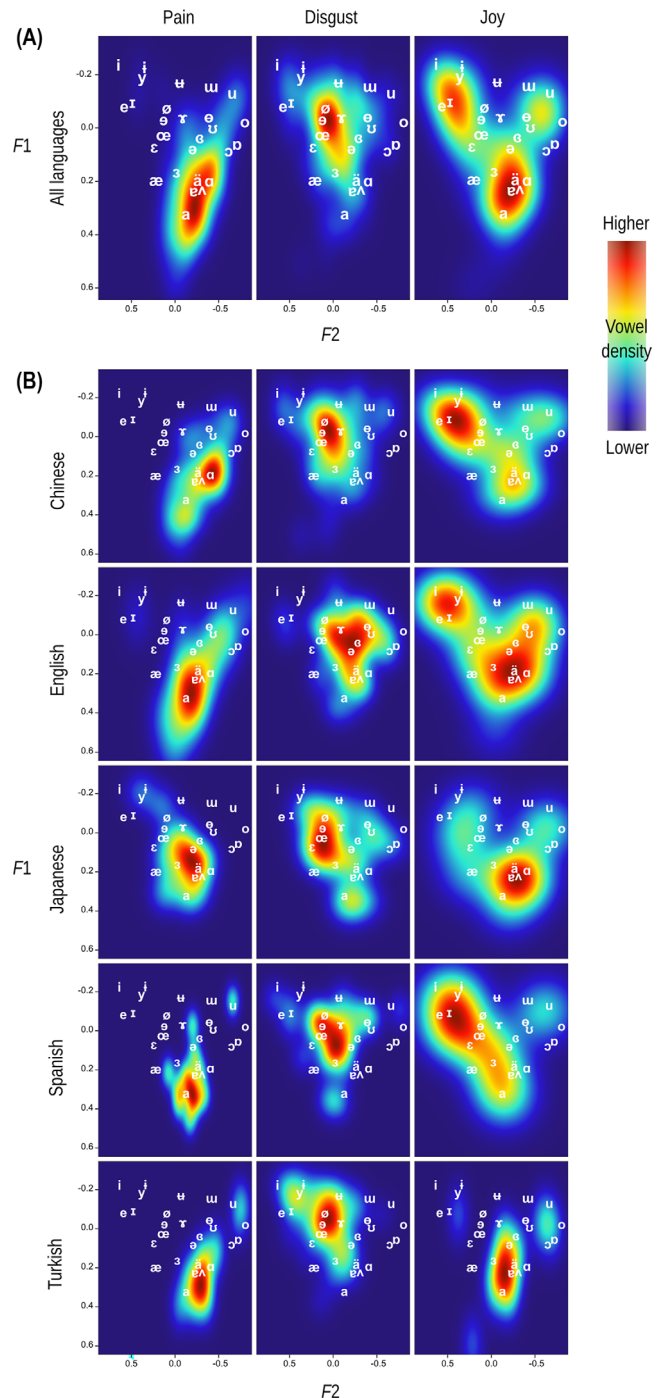

FIG. 4. (Color online) Vowel density maps show different vowel spaces for emotional vocalizations of pain, disgust, and joy that are largely preserved across the five languages. Color gradients show distribution densities for IPA vowels wherein red colors indicate relatively higher densities and blue indicates lower densities. Formant measures $F1$ and $F2$ correspond to the frequencies of the first and second formants for each vowel based on speaker-normalized values, normalized in relation to the neutral schwa vowel (see Methods) based on the one-tube model of the vocal tract (Behrman, 2007). Source code for creating these figures is provided in the supplementary material (SuppPub7.zip).

vocalizations as either aberrant sounds or primary interjections, based on known interjections for each given emotion and language (plus English) derived from dictionaries and discussions with native speakers. At least one coder was a

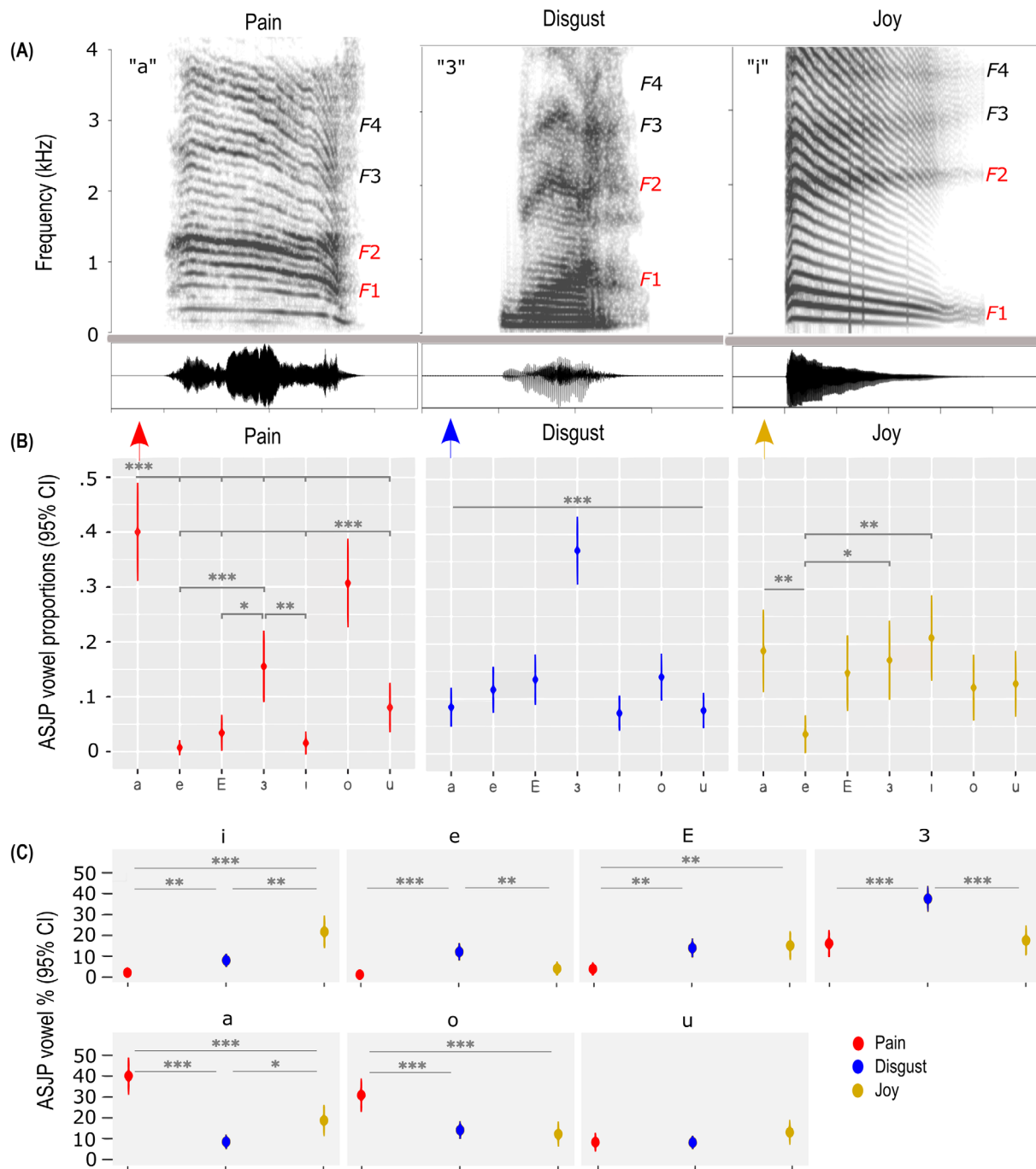J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.* 3129

FIG. 5. (Color online) Multinomial models reveal robust vowel signatures in pain, disgust, and joy vocalizations. (A) example spectrograms of vocalizers producing the most common vowels for each emotional context: a pain vocalization characterized by the vowel "a" with narrow formant $F1$–$F2$ spacing, a disgust vocalization characterized by [ɜ] with moderate $F1$–$F2$ spacing, and a joy vocalization that finishes with the vowel "i," with wide $F1$–$F2$ spacing. Readers may listen to audio recordings of these vocalizations available in the supplementary material, SuppPub10.zip, with captions for these recordings provided in SuppPub3.pdf. Spectrograms were created with a spectral window length of 40 ms and time step of 5 ms; male vocalizers were selected for these examples due to a dense harmonic spectrum. (B) ASJP vowel proportions varied systematically with emotional context. Significant pairwise comparisons between pairs of vowels are indicated by the presence of ticks (*$p < 0.05$,**$p < 0.01$, ***$p < 0.001$) (C) comparing vowels across emotional contexts, we corroborate the results observed in panel B, wherein the low vowels "a" and "o" are most prominent in pain, the central vowel "3" and mid front vowel "e" are most prominent in disgust, and the high front vowel "i" is most prominent in joy. The vowel "u" is uncommon across emotional vocalizations. In both (B) and (C), red corresponds to pain, blue to disgust, and yellow to joy vocalizations. Asterisks indicate *post hoc* comparisons of estimated marginal means derived from multinomial models, *mvt*-adjusted for multiple comparisons, *$p < 0.05$,**$p < 0.01$, ***$p < 0.001$.

native speaker of each given language. Any coding discrepancies were discussed and resolved. A total of 41 vocalizations (containing 77 vowels) were coded as potential interjections and hence omitted from all further analyses

(see dataset SuppPub8, column labelled "inter"). The final sample of formant measures thus derived from 494 vowel segments: 130 from pain vocalizations, 109 from joy, and 255 from disgust vocalizations (see Table S5 in

SuppPub3.pdf for breakdown by vocalizer sex and language). Note that joy and pain vocalizations were much more likely to be extremely high-pitched compared to disgust, hindering formant measurement (see the following: *Formant frequency analysis from vocalizations*), and thus resulting in a relatively higher proportion of disgust vocalizations.

B. Analyses

1. Formant frequency analysis from vocalizations

Vocal stimuli were uploaded as WAV files and edited in *Praat* 6.2.23 (Boersma and Weenink, 2022). Stimuli were first manually quality-checked to ensure a high signal-to-noise ratio and those that contained background noise or clipping which could interfere with acoustic analysis were removed. Formants were measured from 494 vowel segments across all emotional contexts to quantify vowels in nonlinguistic vocalizations. We manually measured the first four formant frequencies ($F1$–$F4$) of each vocalization in the open-source and interactive R package *soundgen* using the *formant_app* function (Anikin, 2019). This software offers an innovative interface in which visual and auditory feedback can be combined to manually track and correct formant contours derived from linear predictive coding (LPC), ensuring relatively robust formant measures (Anikin *et al.*, 2023). Spectrograms were visually inspected to verify the fit of formant tracks to spectral peaks followed by manual adjustment of LPC spectral smoothing and visual inspection of vowel quality. We applied a window length of 50 ms, a time step of 5 ms, and spectral smoothing between -1.5 and -1, as recommended for formant analysis in nonlinguistic vocalizations (Anikin *et al.*, 2023).

Soundgen allows the user to manually select the exact region in a recording from which to measure formants. Formants were measured from the steady state mid-point of each perceptually distinct vowel within each vocalization. This steady state could be a few milliseconds to a few seconds long, depending on the voiced duration of the vowel. If a vocalization contained multiple different vowels, we measured each unique vowel once from its steady state. In the case of gradual formant transitions, such as in diphthongs, we measured the final steady state of each vowel in the pair. From measures of formants $F1$ to $F4$, we computed apparent formant spacing (ΔF) and apparent vocal tract length using the verified regression method to speaker-normalize $F1$ and $F2$ values (Reby and McComb, 2003; Anikin *et al.*, 2023).

Reliable formant measurement requires a relatively dense harmonic structure (Fitch *et al.*, 2024), and thus, vocal segments containing vowels with a fundamental frequency (f_o) exceeding approximately 400 Hz were typically omitted from analyses. This was most common in joy vocalizations (f_o range 92–616 Hz) and pain vocalizations (f_o range 122–643 Hz) that tended to be relatively higher pitched than disgust vocalizations (f_o range 93–492 Hz). We also omitted vowel segments that resembled gagging

without a discernible vowel (most common in disgust) or inhalation through the teeth (most common in pain) due to the lack of clear formant patterns and vowel configuration, as well as closed-mouth vocalizations due to probable nasal formants.

To confirm the reliability of our formant measures, a second researcher measured formants from a stratified subsample of vocalizers for each language, vocalizer sex, and emotion context (three men and three women per language and emotion). This subsample contained 170 annotated vowels from 120 vocalizers, from which 13 vowels were omitted due to aforementioned factors (e.g., high f_o, closed mouth). Inter-rater reliability was extremely high for measurements of the first ($r = 0.97$) and second ($r = 0.97$) formants that largely determine vowel quality and high for formant spacing ($r = 0.86$).

2. Quantifying vowels from formant frequencies

The relative spacing between the first two formants, $F1$ and $F2$, largely determines the vowel quality of any vocal sound (Behrman, 2007). For example, [u] is produced with narrow $F1$–$F2$ spacing, typically achieved by rounding the lips or raising the back of the tongue to constrict the posterior oral cavity. In contrast, [i] is produced with wide $F1$–$F2$ spacing typically achieved by spreading the lips along with pushing the tongue forward and upward (Behrman, 2007). Our $F1$–$F2$ formant measures were thus used to quantify the vowel quality of each nonlinguistic vocalization.

We quantified IPA vowels corresponding to each vowel segment of each vocalization based on their $F1$-$F2$ coordinates (see Fig. 4). Vowel quality was further verified using the *soundgen* vowel space map and audio playback functions (Anikin *et al.*, 2023). Because formant frequencies and their relative spacing are lower in taller individuals with longer vocal tracts (Fitch, 1997; Pisanski *et al.*, 2014, for meta-analysis), formant measures were speaker-normalized based on apparent vocal-tract lengths derived from $F1$–$F4$ (Anikin *et al.*, 2023). We additionally measured and speaker-normalized the first four formants from audio recordings of IPA vowels openly available online (Internationalphoneticalphabet.org and Wikipedia.org), averaging the measures from both audio sets to derive vowel spaces (see Fig. 4). IPA vowels were then re-coded to correspond to the broader ASJP vowel categories, as we did with interjections (see Table S2 in SuppPub3), and these ASJP vowels were used in multinomial regression models.

3. Statistical analysis of nonlinguistic vocalizations

All code, datasets and analyses are provided as supplementary material in SuppPub7.zip, SuppPub8.ods and SuppPub9.HTML respectively. Data were analyzed with multinomial regression models using the *mblogit()* function from the R package mclogit (Elff, 2022) to test whether the ASJP vowels characterizing nonlinguistic vocalizations varied for expressions of pain, disgust, and joy. The ASJP vowels, based on formant analyses of 494 vowel segments

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.* 3131

excluding potential interjections, were dependent variables in all models, and emotional context was included as the fixed factor. We compared several models with or without the random effect of vocalizer identity (IDvocalizer) and the additional fixed effect of vocalizer language (Language) to determine the best-fitting model. Based on the deviance and AIC of all models (see supplementary material, SuppPub9. HTML), the best fit was achieved with model 4, which included both emotional Context and Language as fixed effects, and no random effects. In addition, we re-ran this model including an interaction between Context and Language to test whether differences in vowels across emotions were observed across languages. Full model outputs including analyses of variance are reported in the supplementary material (SuppPub9.HTML). All pairwise comparisons of estimated marginal means derived from the models were adjusted for multiple comparisons using a multivariate t distribution (*mvt* adjustment) with the same covariance structure as the estimates.

C. Results and discussion

Figure 4 plots the vowel spaces for pain, disgust, and joy vocalizations based on our speaker-normalized $F1$-$F2$ formant measures (see Methods). These density plots of IPA vowels effectively illustrate differences in the distributions of vowel usage across the three emotional contexts [Fig. 4(A)], with similar vowel patterns observed for emotional contexts across languages, especially for pain and, to a lesser extent, for disgust [Fig. 4(B)]. We plot IPA vowels here for illustrative and comparative purposes, but our models were conducted on ASJP vowels, as they were for interjections. Thus, the IPA vowel patterns observed in Fig. 4 were corroborated with multinomial regression models that showed a strong significant effect of emotional context on ASJP vowel quality, confirming that vowels differed significantly across pain, disgust, and joy vocalizations collapsing across languages ($\chi^2 = 136.05$, $p < 0.001$, SuppPub9.HTML).

Figure 5 shows the results of pairwise comparisons derived from the multinomial models, comparing ASJP vowels within each emotion context [Fig. 5(B)] and between emotions for each vowel [Fig. 5(C)], adjusted for multiple comparisons (*mvt* adjustment). Clear and distinct vowel signatures were observed for nonlinguistic vocalizations expressing each emotion. First, pain vocalizations showed a very high proportion of "a" vowels [40% ± 0.05 standard error of the mean (sem), 95% confidence interval (CI): 0.31–0.49] and "o" vowels (31% ± 0.04 sem, 95% CI: 0.23–0.39). These are low and back vowels typically produced with a wide-open mouth or rounded lips, respectively, resulting in narrow spacing between $F1$ and $F2$, as illustrated in the left-most spectrogram in Fig. 5(A). The proportion of "a" vowels in pain vocalizations was significantly higher than the proportions of all other vowels produced when communicating pain (all $Z > 3.70$, all $p < 0.01$), except "o," which was the second most

prominent vowel in pain with a significantly higher proportion compared to all other vowels [except [a] and [ɜ], all $Z > 4.47$, all $p < 0.01$; see Fig. 5(B)]. Vowel comparisons among (rather than within) emotions further showed that "a" and "o" vowels were significantly more prominent in pain than in disgust or joy vocalizations [Fig. 5(C); all $Z > 3.68$, all $p < 0.001$].

Second, disgust vocalizations were characterized by a very high proportion of "3" vowels (37% ± 0.03 sem, 95% CI: 0.31–0.43). These schwa-like central vowels are produced with little articulation and are characterized by moderate spacing between $F1$ and $F2$, as in the central spectrogram in Fig. 5(A). The proportion of "3" vowels in disgust vocalizations was significantly higher than the proportions of all other vowels in disgust [all $Z > 5.2$, all $p < 0.0001$; Fig. 5(B)]. This was further corroborated by comparisons across emotions, showing that "3" vowels were much more prevalent in disgust than in pain or joy vocalizations [all $Z > 4.2$, all $p < 0.0001$; Fig. 5(C)].

Third, the most common vowel observed in joy vocalizations was "i" (21% ± 0.04 sem, 95% CI: 0.13–0.29). These are high front vowels that are often produced with spread lips (akin to a smile) and thus characterized by a wide spacing between $F1$ and $F2$, as in the right-most spectrogram illustrated in Fig. 5(A). Joy vocalizations had significantly more "i" vowels compared to pain and disgust vocalizations [all $Z > 3.25$, all $p < 0.01$; Fig. 5(C)]. Joy vocalizations were also characterized by a high proportion of "a" vowels [19% ± 0.04 sem, 95% CI: 0.11–0.26; Fig. 5(C)], much less than what we observed for pain ($Z = 3.7$, all $p < 0.001$), but significantly more than in disgust vocalizations [$Z = 2.45$, $p = 0.037$; Fig. 5(C)].

In our multinomial models, we additionally tested for an interaction between emotional context and language to examine whether the variance we observed in vowels across emotions was present in most languages rather than being driven by a subset of languages (see supplementary material, SuppPub9.HTML). The interaction effect between context and language was significant, indicating some cross-linguistic variance ($\chi^2 = 90.48$, $p < 0.001$). However, overall, the models confirmed that the key vowel signatures we observed in pain, disgust, and joy vocalizations described previously were indeed characteristic of most languages, despite minor deviations, as also illustrated in the country-level vowel density maps [Fig. 4(B)]. For example, in all five languages, "a" or "o" vowels were the most common in pain vocalizations followed by "3," whereas "3" was the most common vowel in disgust vocalizations across all five languages (see Fig. S4 in SuppPub3.pdf). Joy vocalizations showed a relatively high proportion of "i" vowels in Chinese, English, and Spanish, albeit not in Japanese and Turkish. While these cross-linguistic trends are reassuring, the large number of comparisons and relatively small samples of vocalizations within countries meant that our interaction analyses were statistically underpowered and should be interpreted with caution.

3132   J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.*

Our formant analyses of nonlinguistic vocalizations thus showed a distinctive vowel signature for each emotional vocalization: pain vocalizations were mainly characterized by "a" and "o" vowels, disgust by schwa-like central vowels "3", and joy vocalizations by "i" and "a" vowels. These patterns align strongly with what we observed for pain interjections, and moderately with what we observed for joy interjections, however, the high proportion of central vowels in disgust was specific to nonlinguistic vocalizations and not shared by interjections. In the next section, we conjecture about the potential mechanisms and implications of these comparative findings.

IV. GENERAL DISCUSSION

Having examined the forms of expressive interjections and emotional nonlinguistic vocalizations in the previous sections, here we bring our observations together to answer the questions at the core of this study. First, do interjections expressing pain, disgust, and joy show resemblances across the world's languages? Our analyses suggest that it is indeed clearly the case for pain interjections, which show a remarkable cross-linguistic convergence. Joy and disgust interjections, on the other hand, show no meaningful convergence or divergence despite very small incidental differences (Sec. II C 1). Second, are certain vowels more prevalent in interjections across languages, and if so, which vowels? We show that pain interjections are largely characterized by a preference for "a" vowels and wide falling diphthongs. To some extent, the same vowels are also frequent in joy interjections across regions, while disgust interjections exhibit few consistent vowel signatures, and these vary from one region of the world to another (Sec. II C 2). Third, can the vocalic tendencies observed in interjections be ostensibly traced back to nonlinguistic vocalizations expressing the same emotions or affective states? Our formant analyses reveal that vowel patterns in nonlinguistic vocalizations are shared across languages for each emotional context, with "a," "3," and "i" being most frequent in pain, disgust, and joy vocalizations respectively (Sec. III C). In interjections, the same vowel patterns are only robustly observed for pain.

Except for pain, our results do not offer strong evidence that nonlinguistic vocalizations influence the vowel components of interjections expressing joy or disgust. In this respect, pain interjections clearly stand out for their strong resemblance with nonlinguistic vocalizations in vowel patterns. Interestingly, based on our results, there is reason to further hypothesize that the forms of pain interjections may actually influence the forms of joy interjections, and in this way, that interjections for further emotional categories may be shaped to some extent by nonlinguistic vocalizations – with pain vocalizations appearing to play a pivotal role. While this intriguing hypothesis remains to be tested, we suggest that such an influence would be indirect, via well-traveled semantic paths between pain and other emotive interjections, as discussed further in Sec. IV B.

A. Comparing interjections to nonlinguistic vocalizations

Few studies to date have examined the vowel space of nonlinguistic vocalizations, with one recent study showing that human vocalizations on average contain a high proportion of a-like vowels (Anikin *et al.*, 2023). However, as noted in the introduction, there is good reason to predict that the relative proportions of different vowels will vary depending on the emotion being expressed by a given vocalization. This could be due to a number of factors such as the physical mode of vocal production (e.g., degree of mouth opening; Anikin *et al.*, 2024), sound-symbolic associations (e.g., with "bright" vowels related to positive valence; Butcher, 1974), and/or evolved form-function associations wherein the vocal sounds of humans, like other animals, appear to be shaped by natural and sexual selection to maximize or exaggerate the expression of certain traits and states (Darwin, 1872; Morton, 1977; Ohala, 1984; Pisanski *et al.*, 2022, for review).

As presented in Sec. II, we show that pain interjections have distinctive vowel signatures that are broadly consistent across different languages and regions of the world. Namely, open vowels ("a" in ASJP) and wide falling diphthongs ("ai," "a3," "au," "ay," and "aw" in ASJP) are significantly more prevalent in pain interjections than in the rest of the length-matched lexicon. This aligns with the vowel signature we found for nonlinguistic pain vocalizations across five languages, which also feature significantly more "a" vowels than any other vowels. Importantly, these "a" vowels were also significantly more common in pain vocalizations than in disgust or joy vocalizations. These results align with the predictions of Dingemanse (2023) and with our prediction that vocal sounds intended to express pain are likely to include more low, central vowels produced with a wide-open mouth, as opening the mouth to vocalize is a common reflex when experiencing physical pain (Helmer *et al.*, 2020). Thus, volitional pain vocalizations and linguistic interjections may to some degree be "iconic" conventionalizations of reflexive pain vocalizations. Here, we may note that if one chose to embrace the Peirce (1955) classical distinction between "icon" and "index," pain interjections may also be candidate "indices," a status they would then share with nonlinguistic vocalizations. That is, pain interjections would not resemble nonlinguistic vocalizations because they imitate them (Peirce's iconicity) but because they result from the same physiological or functional constraint as nonlinguistic vocalizations (Peirce's notion of index).

We also observed a high proportion of wide falling diphthongs in pain interjections, that is diphthongs starting with "a" such as "ai" (pronounced in English as "Ayy!") and "aw" (pronounced as in "Ouch!"). As we did not predict this feature *a priori*, and it only became apparent as we began to examine our linguistic data, our methodology to

J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.* 3133

16 November 2024 15:11:14

assess vowels in nonlinguistic vocalizations did not test for diphthongs. Therefore, whether wide falling diphthongs represent another point of convergence between interjections and nonlinguistic vocalizations for pain remains an open question for future research. In the absence of a clear form-function hypothesis that would justify the over-representation specifically of *diphthongs* in nonlinguistic pain vocalizations, we have no particular reason to hypothesize that their presence in interjections should be traced back to nonlinguistic vocalizations. Yet there is some support for an influence of nonlinguistic vocalizations on pain interjection vowels—and importantly, what is observed with pain interjections strikingly contrasts with disgust and joy, where our results appear to rule out such an influence, as discussed in the following.

Our results further indicate that nonlinguistic vocalizations intended to express disgust exhibit a clear preference for central vowels "3," well distinct from pain and joy. This aligns with our prediction that vocal sounds intended to express disgust will have a higher proportion of central schwa-like vowels produced without much articulation, that is without much rounding of the lips or manipulation of the jaw and tongue, reflecting a reflexive oral response of gagging or expelling food from the mouth. This suggests that volitional disgust vocalizations, like pain vocalizations, may also be iconic extensions of reflexive disgust vocalizations. However, this prevalence of central vowels was not matched in disgust interjections. In fact, this absence of systematic vowel patterns suggests that disgust interjections may be largely arbitrary in this respect.

As for joy, the vowel signature of nonlinguistic vocalizations was not as strong as that of pain or disgust, yet we observed a preference for "i" followed by "a," the latter of which aligns with our predictions. However, here too, this pattern remains generally unmatched in joy interjections, which share some features with joy vocalizations ("a") and with pain interjections ("a" and diphthongs), but no obvious preference for "i." Therefore, disgust and joy do seem to imprint distinctive vowel signatures onto nonlinguistic vocalizations, but these signatures are not robustly passed onto linguistic interjections.

While our data, which focused on vowels, cannot speak to the universality of emotional vocal signals more broadly, our results support the hypothesis that emotional vocal utterances share a vowel signature across disparate languages and regions of the world, and that this may be rooted in their shared function. Cross-cultural research on nonverbal vocalizations is on the rise (Brooks *et al.*, 2023; Ćwiek *et al.*, 2021; Laukka and Elfenbein, 2021; Kamiloğlu *et al.*, 2020; Sauter *et al.*, 2010), but much more work is needed to assess the extent to which vocalizations such as laughter, crying, and screaming share some universal acoustic signatures, and which socioeconomic, geographic, cultural, or phylogenetic factors may explain variance in their production and use.

B. Interjections influence each other: The role of pain?

With respect to joy interjections, our data point to the hypothesis that their forms may be directly influenced not by nonlinguistic vocalizations, but by the form of pain interjections (and so perhaps indirectly by pain vocalizations).

In three regions out of five, joy interjections exhibit a preference for wide falling diphthongs, like the one observed with pain interjections; "a" vowels are also prevalent in joy interjections in the Asian languages of our sample. This resemblance between pain and joy interjections may be the consequence of diachronic processes. Specifically, it could result from the fact that a number of joy interjections may, in the past, have been pain interjections. This is supported by the "colexifications" observed in our data.

Linguists talk about *colexification* when a word has several meanings (François, 2008). If a word means both *a* and *b*, we say that it *colexifies a* and *b*; and that *a* and *b* are *colexified* by the word in question. Some interjections in our data colexify pain and joy (17 tokens, or just under 7% of all joy interjections), and some colexify pain and disgust (nine tokens, or just above 7% of all disgust interjections). A smaller proportion of these subsets colexify pain, disgust, and joy altogether (three tokens across the dataset). The number of pain/joy colexifications is too low to explain the formal statistical resemblances between these interjections, but data reviewed in the following suggest that these colexifications are just the "tip of the iceberg," so to speak. They likely indicate that a larger number of interjections in our data may have historically evolved from one emotion category to the other.

Research on semantic change (i.e., the way the meanings of words evolve) has shown that in all languages in the world, words frequently change meaning. It is also established that they do so following relatively regular "semantic paths" (Vanhove, 2008; Juvonen and Koptjevskaja-Tamm, 2016). For instance, we know that across the world's languages, the meaning "(piece of) rock" often gives rise to the meaning "seed," so that words that mean "rock" can often gain the meaning "seed." That is, "rock-to-seed" is a well-traveled "semantic path" (see the Catalogue of Semantic Shifts, Zalizniak, 2020). In addition, we know that before a word can shift from sense *a* to sense *b*, the word often goes through a phase where it means both *a* and *b* simultaneously for some time (Evans and Wilkins 2000). That is, a word meaning "rock" might gain the meaning "seed" and mean both "rock" and "seed" for a while; later it might lose the meaning "rock," so that it will simply mean "seed." In other words, when a word colexifies two meanings *a* and *b*, often it will end up traveling the whole semantic path from *a* to *b*. Therefore, if a number of words in a dataset colexify *a* and *b*, this suggests "*a*-to-*b*" may be a well-traveled semantic path. This implies that many other words currently meaning *b* probably used to mean *a* at some point in the past, but have been traveling the said semantic path, and have

3134   J. Acoust. Soc. Am. **156** (5), November 2024

Ponsonnet *et al.*

16 November 2024 15:11:14

arrived. On the contrary, if *a* and *b* are not found to be colexified, this suggests that words with the meaning *a* do not regularly change their meaning to *b*.[9] Overall, statistically, the more frequent a colexification between *a* and *b*, the more likely it is that words meaning *b* used to mean *a*.[10]

We can now apply this scenario to our interjection dataset. As mentioned previously, in our data, 17 interjections colexify pain and joy (7% of joy interjections). It is therefore plausible that a larger proportion of the joy and pain interjections in our data set *used* to express pain in the past, and have now traveled the whole way from expressing pain to expressing joy. Such a path offers a hypothetical scenario to explain the resemblances in vowel signatures between pain and joy interjections in our data.

With this scenario in mind, as the proportions of pain-disgust and pain-joy colexifications are equivalent in our dataset (also 7% of all disgust interjections), we may wonder why we observe resemblances between joy and pain interjections, and not so much between joy and disgust interjections. Looking at the other senses colexified by these interjections helps answer this question. The most frequent colexification of pain interjections is surprise (e.g., *yekaye* in Katetye, Australia). This is also the most frequent colexification of joy interjections (e.g., *ina* in Guahibo, Columbia, and Venezuela), and threefold colexifications between pain, surprise and joy are attested (e.g., *aragóy* in Hiligaynon, Philippines). This points to "pain-surprise-joy" as another possible semantic path linking pain and joy, and the availability of this second semantic path suggests that an even greater proportion of joy interjections may have expressed pain at some point in time. Colexifications of disgust with surprise are rarer, and no such clear alternative semantic path is available between pain and disgust. These hypotheses offer testable predictions for future studies investigating the origins of emotive interjections.

C. Limitations and future directions

This study was conceived of and designed as an exploratory investigation of an intuitive yet complex question that had not yet been thoroughly empirically tackled. It is therefore important to outline the limitations of the present work alongside promising directions for future research.

First and foremost, in this study, we only examined three emotion categories. In future work, researchers will need to consider a broader variety of affective states to get a fuller picture of universal vowel patterns in interjections and their potential links to nonlinguistic counterparts, and to one another. In addition, our volitional nonlinguistic vocalizations were produced in response to specific contexts (e.g., burning for pain, achievement for joy). Using specific contexts has its benefits, allowing us to standardize the affective experience (emotional intensity, valence, etc.) much more than if we were simply to ask participants to produce a general "pain" or "joy" vocalization, which could be interpreted in myriad ways. However, these contexts can give rise to specific acoustic forms that may differ from those in

vocalizations expressing the same emotion, but in another context or of another intensity (e.g., giving birth for pain). Indeed, pain vocalizations in humans representing mild, moderate, or severe pain differ in their spectrotemporal structures with more intense pain linked to a higher proportion of nonlinear acoustic phenomena and thus more vocal harshness in both volitional and reflexive pain vocalizations (Koutseff *et al.*, 2018; Raine *et al.*, 2019; Valente *et al.*, 2025). It is possible that pain vocalizations of variable arousal also show variable vowel patterns, such as a relatively higher proportion of "a" vowels in extreme versus mild pain (potentially linked to the degree of mouth opening, see, e.g., Anikin *et al.*, 2024), though this has yet to be tested. Thus, we propose that not only should future studies include a broader range of emotions, but also a broader range of contexts and arousal levels within each emotion category. We also encourage replication studies to include a broader range of cultures including vocalizations from non-WEIRD (not Western, Industrialized, Educated, Rich, and Democratic) populations (Henrich *et al.*, 2010), particularly those without access to the Internet and popular media portrayals of emotional expressions.

Our rationale for using volitional rather than reflexive or "spontaneous" vocalizations was chiefly that this allowed us to standardize the exact context in which vocalizations were produced. Given the aforementioned variance within emotion categories as a function of arousal, this was an important control. This decision was also motivated by constraints on formant measurement (Fitch *et al.*, 2024): volitional vocalizations are usually relatively lower in amplitude and thus lower in fundamental frequency than their spontaneous counterparts, allowing for the retention of a higher proportion of vocalizations for formant analysis. Nevertheless, using volitional vocalizations introduces a few key shortcomings. First, volitional vocalizations are by definition produced voluntarily and may be more stereotyped than their spontaneous counterparts (Anikin and Lima, 2018). Indeed, recent evidence suggests that volitional vocalizations in humans may be learned through vocal production learning, much like speech (Pisanski *et al.*, 2024). Second, even though vocalizers were specifically instructed not to produce interjections, and we removed vocalizations featuring interjections from our data, in principle, there is a possibility that the vocalizations that participants produced in our study were, to some extent, influenced by their knowledge of interjections lexicalized in their native languages. This potentially introduces some degree of circularity in our argument. This risk seems however limited, as for two out of three emotional contexts, we do *not* find a robust resemblance between interjections and nonlinguistic vocalizations. In future work, we can avoid this circularity altogether by examining spontaneous (i.e., less voluntary) vocalizations for a broad range of real-life emotional contexts, with the unavoidable caveat that high-arousal vocalizations will be higher-pitched and less amenable to formant analysis. One way around this may be to analyze vowels in vocalizations using perceptual

vowel-discrimination tasks in samples of human listeners, assuming that vowels can often be discriminated even when voice pitch is too high to actually visualize or measure formants on a spectrogram. These will surely be more subjective than formant-based vowel measurements but will allow us to assess vowel quality even in very high-pitched vocalizations.

In our study, vowels were based on text transcriptions for interjections, and we only had audio recordings for vocalizations. In future work, we endeavor to test for broader acoustic signatures in audio recordings of both vocalizations and interjections. Research has shown that spectrotemporal parameters of vocal signals including fundamental frequency (pitch), duration, loudness, and nonlinear phenomena (which give vocalizations a rough and harsh quality) tend to follow predictable form-function mappings, including vocalizations of pain that are often very harsh (Koutseff *et al.*, 2018; Raine *et al.*, 2019; Pisanski *et al.*, 2022; Valente *et al.*, 2025). Our aim is to test whether these broader acoustic signatures are shared between vocalizations and interjections sharing the same communicative function.

As another limitation, in both vocalizations and interjections, we did not examine consonants or other properties such as syllable structure or length, though length is implicitly accounted for in our statistical analyses by comparing interjections to length-matched lexical samples. Given the size of our dataset, vowels offered a stronger statistical basis because both vocalizations and interjections contain more vowels than consonants (see Sec. II C 2). Consonant signatures and other properties (particularly syllable structure, see Vallery and Lemmens, 2021) will be examined once we have gained access to more interjection data. As mentioned in Sec. II A, collecting reliable and precise data on interjections in the world's languages is a challenge because of the lack of homogeneity in the way interjections appear in grammars and dictionaries if they appear at all. In future studies, we intend to expand the dataset to languages that are also represented in the Lexibank dataset, while maintaining (and enhancing) the geographical and phylogenetic diversity in the sample, with the additional benefit of having these effects explicitly controlled in the statistical analysis.

This will also help resolve the differences between the analyses performed with an ASJP or Lexibank lexical baseline, which may arise from differences in language coverage. However, they may also underscore that the small ASJP lexicons may be insufficiently robust to estimate each language's phonological envelope (the frequency statistics of the phonemes of a language), leading to noisy estimations of the cross-language distances. For this reason, the results discussed in this paper are those for which both ASJP-based and Lexibank-based analyses agree, to benefit from both the breadth of ASJP (in number of languages) and the depth of Lexibank lexicons (in number of words per lexicon) and to reach a reasonable level of confidence.

V. CONCLUSION AND PERSPECTIVES

To conclude, these preliminary comparative analyses provide fascinating clues into potentially universal and iconic vowel signatures in emotive interjections across the world's languages and their possible origins. In future research, we endeavor to test and explore the results of this study further by targeting its limitations—particularly, expanding the range of emotions, investigating other properties of interjections such as their consonant makeup, syllable structure, and prosodic contours, and acoustically comparing interjections with spontaneous rather than volitional vocalizations by analyzing voice recordings of both interjections and vocalizations. Importantly, our observations in this study also raise the question of why some features of vocalizations associated with certain emotional contexts, like pain, "make it" into conventionalized languages, while for some other emotional contexts, this does not seem to happen. Is this due to how effectively certain low-level acoustic features of nonlinguistic vocalizations can be preserved in speech, to their social communicative functions (e.g., cathartic vs communicative), or to the different adaptive functions of emotional expressions in human life and communication? While our study has not yet brought definite answers to these questions, we hope that it has shed light on their importance and paved the way for further research.

SUPPLEMENTARY MATERIAL

See the supplementary material for the complete dataset of interjections, references for the sources of data, an interactive map of the languages, a description of ASJPcode, details about the participants in the study on nonlinguistic vocalizations, sound recordings of vocalizations, additional descriptive statistics, the complete pipeline of all the statistical analyses, and code.

ACKNOWLEDGMENTS

AUTHOR DECLARATION
Conflict of Interest

The authors declare they have no conflicts to disclose.

Ethics Approval

Ethical approval for the acoustic recording of human subjects and analysis of their nonlinguistic vocalizations was provided by the Comité d'Ethique du CHU de Saint-Etienne (IRBN692019/CHUSTE) for participants recorded via the Prolific platform, and by the Ethical Review Committee for Experimental Research involving Human Subjects at the Graduate School of Arts and Sciences and the College of Arts and Sciences, The University of Tokyo, for participants recorded in Japan (# 962). Informed and written consent was obtained from all participants.

DATA AVAILABILITY

Data and codes developed for data processing and analysis are available as supplementary material. Datasets are also openly available on GitHub (https://github.com/keruiduo/SupplMat_JASA_2024).

---

[1]See Sec. II A 2 on transcription conventions.

[2]There are 132 languages represented in Fig. 1(A). Four are analyzed in both the interjection and vocalization studies, while 127 are only analyzed in the interjection dataset and one only in the vocalization dataset. An interactive version of the map with links to each language's Glottolog record is provided as SuppPub1.html.

[3]We also checked what the sources listed under "exclamation," "onomatopoeia," and "particle." Some linguists use the label "exclamation" for what we call "interjection" here. Onomatopoeia and particles, on the other hand, differ more clearly from interjections: the former imitate events, like *bang* in English imitates a shock or an explosion; the label "particle" is versatile, and tends to designate somewhat unclassifiable words. In this study, we look at interjections to the exclusion of "onomatopoeia" and "particles." That is, words imitating someone yelling in pain, like *waaah*, were excluded. However, searching for "onomatopoeia" and "particle" throughout our sources revealed a significant number of interjections mistakenly tagged as one or the other.

[4]Such interjections, which linguists call "secondary" interjections, are very widespread (e.g., *shit* in English). As it is often difficult to decide where an interjection is identical to a word and where it differs, attempting to exclude secondary interjections from our dataset did not make sense. In addition, one may hypothesize that acoustic properties influence speakers' choices to use certain words rather than others as interjections.

[5]The subsampling procedure implemented to control for the dataset imbalance results in the non-independence of the subsamples. It prevents us from using a simple paired t-test or its non-parametric equivalent, the Mann-Whitney U test.

[6]The Damerau-Levenshtein distance differs from the Levenshtein distance by including transpositions (swaps) among the allowable operations. It is consequently more adequate because metathesis (the swap of two adjacent phonemes in a sequence) is irrelevant for our research questions, which deal with phoneme composition rather than ordering. Matching the length between interjections and lexical items is not a requirement of the Damerau-Levenshtein algorithm but it ensures comparability.

[7]The p-value was defined as the proportion of samples with an average distance more distant from the mean of the distribution than the average distance for the interjections (i.e., the proportion of samples more "extreme" than interjections, considering a bilateral test). The effect size was computed as the Cohen's d, i.e., the distance between the mean of the distribution and the average distance for the interjections, divided by the standard deviation of the distribution.

[8]Frequency corresponds here to how frequently a target vowel appears in the lexical forms. If a form consists of six phonemes and the vowel appears twice, then the frequency for this form is $2/6 = 1/3$. For wide falling diphthongs, the frequency is the number of occurrences of these diphthongs divided by twice the number of phonemes. Average frequencies can be computed for languages or regions of the world, both for interjections and lexicons.

[9]Of course, certain forms may maintain both senses *a* and *b*, and never lose meaning *a*. Nevertheless, the colexification of two meanings in a single form *conditions* the shift of this form from one meaning to the other. Therefore, meanings that attest colexifications must, overall and cross-linguistically, lead to one another more often than those that do not attest colexifications.

[10]As pointed out by an anonymous reviewer, senses *a* and *b* do not have to be distinct in the original word for the mechanism to apply, and this is particularly plausible for interjections. For instance, we can imagine an interjection that applies in contexts where one experiences pain and surprise at the same time (like when knocking one's head against the corner of a shelf for instance)—here the senses *a* and *b* may always have been merged in the same interjection. However, in time, this interjection may specialize to mean just surprise. Irrespective of whether a form starts as *a* or as *a* and *b*, the existence of colexifications between *a* and *b* indicates that statistically, there are chances that forms meaning *b* used to (also) mean *a*.

Ameka, F. (**1992**). "Interjections: The universal yet neglected part of speech," J. Pragmatics **18**(2), 101–118.

Anikin, A. (**2019**). "Soundgen: An open-source tool for synthesizing non-verbal vocalizations," Behav. Res. **51**(2), 778–792.

Anikin, A., Bååth, R., and Persson, T. (**2018**). "Human non-linguistic vocal repertoire: Call types and their meaning," J. Nonverbal Behav. **42**, 53–80.

Anikin, A., Barreda, S., and Reby, D. (**2023**). "A practical guide to calculating vocal tract length and scale-invariant formant patterns," Behav. Res. **56**, 5588–5604.

Anikin, A., Canessa-Pollard, V., Pisanski, K., Massenet, M., and Reby, D. (**2023**). "Beyond speech: Exploring diversity in the human voice," iScience **26**(11), 108204.

Anikin, A., and Lima, C. (**2018**). "Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations," Q. J. Exp. Psychol. **71**(3), 622–641.

Anikin, A., Pisanski, K., and Reby, D. (**2020**). "Do nonlinear vocal phenomena signal negative valence or high emotion intensity?," R. Soc. Open Sci. **7**(12), 201306.

Anikin, A., Valente, D., Pisanski, K., Cornec, C., Bryant, G. A., and Reby, D. (**2024**). "The role of loudness in vocal intimidation," J. Exp. Psychol.: General **153**(2), 511–530.

Behrman, A. (**2007**). *Speech and Voice Science* (Plural Publishing, Inc., San Diego, CA).

Bergen, B. K. (**2004**). "The psychological reality of phonaesthemes," Language **80**(2), 290–311.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (**2016**). "Sound-meaning association biases evidenced across thousands of languages," Proc. Natl. Acad. Sci. U.S.A. **113**(39), 10818–10823.

Boersma, P., and Weenink, D. (**2022**). "Praat: Doing phonetics by computer (6.2.23) [computer software]," https://www.fon.hum.uva.nl/praat/ (Last viewed March 1, 2024).

Brooks, J. A., Tzirakis, P., Baird, A., Kim, L., Opara, M., Fang, X., Keltner, D., Monroy, M., Corona, R., Metrick, J., and Cowen, A. S. (**2023**). "Deep learning reveals what vocal bursts express in different cultures," Nat. Hum. Behav. **7**(2), 240–250.

Brown, C. H., Holman, E. W., and Wichmann, S. (**2013**). "Sound correspondences in the world's languages," Language **89**(1), 4–29.

Brown, C. H., Holman, E. W., Wichmann, S., and Velupillai, V. (**2008**). "Automated classification of the world's languages: A description of the method and preliminary results," Lang. Typol. Universals **61**(4), 285–308.

Butcher, A. (**1974**). "'Brightness,' 'Darkness' and the dimensionality of vowel perception," J. Phon. **2**(2), 153–160.

Charlton, B. D., Pisanski, K., Raine, J., and Reby, D. (**2020**). "Coding of static information in terrestrial mammal vocal signals," in *Coding*

16 November 2024 15:11:14

Strategies Vertebrate Acoustic Communication (Springer, New York), pp. 115–136.

Cohen, J. (**1969**). *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York).

Colombat, B., and Lahaussois, A. (**2019**). "Histoire des parties du discours" ("History of the parts of speech"), Ser. Orbis Supplementa **46**, 441–474.

Cowen, A. S., Elfenbein, H. A., Laukka, P., and Keltner, D. (**2019**). "Mapping 24 emotions conveyed by brief human vocalization," Am. Psychol. **74**(6), 698–712.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., Zeller, J., Winter, B., and Perlman, M. (**2021**). "Novel vocalizations are understood across cultures," Sci. Rep. **11**(1), 10108.

Darwin, C. (**1872**). *The Expression of Emotions in Man and Animals* (Chicago University Press, Chicago, IL).

Damerau, F. J. (**1964**). "A technique for computer detection and correction of spelling errors," Commun. ACM **7**(3), 171–176.

De Carolis, L., Marsico, E., and Coupé, C. (**2017**). "Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features," Language Commun. **54**, 21–35.

Dingemanse, M. (**2011**). "The meaning and use of ideophones in Siwu," Ph.D. thesis, Radboud University Nijmegen, Nijmegen, the Netherlands.

Dingemanse, M. (**2017**). "On the margins of language: Ideophones, interjections and dependencies in linguistic theory," in *Dependencies in Language*, edited by N. J. Enfield (Language Science Press, Berlin, Germany), pp. 195–203.

Dingemanse, M. (**2023**). "Interjections," in *The Oxford Handbook of Word Classes*, edited by E. van Lier (Oxford University Press, Oxford, UK).

Dingemanse, M. (**2024**). "Interjections at the heart of language," Annu. Rev. Linguist. **10**(1), 257–277.

Elff, M. (**2022**). "mclogit: Multinomial logit models, with or without random effects or overdispersion [computer software]," https://cran.r-project.org/package=mclogit (Last viewed March 1, 2024).

Ehlich, K. (**1985**). "The language of pain," Theor. Med. Bioeth. **6**, 177–187.

Erben Johansson, N., Anikin, A., Carling, G., and Holmer, A. (**2020**). "The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features," Ling. Typol. **24**(2), 253–310.

Evans, N., and Wilkins, D. (**2000**). "In the mind's ear: The semantic extensions of perception verbs in Australian languages," Language **76**(3), 546–592.

Feldman Barrett, L. (**2009**). "Variety is the spice of life: A psychological construction approach to understanding variability in emotion," Cogn. Emotion **23**(7), 1284–1306.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., and König, P. (**2016**). "LabVanced: A Unified JavaScript Framework for Online Studies," in *Proceedings of the 2017 International Conference on Computational Social Science IC2S2*, July 10–13, Cologne, Germany.

Fischer-Jørgensen, E. (**1968**). "Perceptual dimensions of vowels," STUF Lang. Typol. Universals **21**(1–6), 94–98.

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am **102**(2), 1213–1222.

Fitch, W. T., Anikin, A., Pisanski, K., Valente, D., and Reby, D. (**2024**). "Formant analysis of vertebrate vocalizations: Achievements, pitfalls & promises," BMC Biol. (in press).

François, A. (**2008**). "Semantic maps and the typology of colexification: Intertwining polysemous networks across languages," in *From Polysemy to Semantic Change*, edited by M. Vanhove (John Benjamins, Amsterdam, the Netherlands), pp. 163–215.

Haiman, J. (**2018**). *Ideophones and the Evolution of Language* (Cambridge University Press, Cambridge, UK).

Hammarström, H., Forkel, R., Haspelmath, M., and Bank, S. (**2024**). "Glottolog 5.0," https://doi.org/10.5281/zenodo.10804357 (Last viewed June 13, 2024).

Harré, R. (**1986**). *The Social Construction of Emotions* (Basil Backwell, New York).

Helmer, L. M., Weijenberg, R. A., de Vries, R., Achterberg, W. P., Lautenbacher, S., Sampson, E. L., and Lobbezoo, F. (**2020**). "Crying out

in pain—A systematic review into the validity of vocalization as an indicator for pain," Eur. J. Pain **24**(9), 1703–1715.

Henrich, J., Heine, S. J., and Norenzayan, A. (**2010**). "The weirdest people in the world?," Behav. Brain Sci. **33**(2–3), 61–83.

Hinton, L., Nichols, J., and Ohala, J. J. (**2006**). *Sound Symbolism* (Cambridge University Press, Cambridge, UK).

Jones, D. (**1954**). "Falling and rising diphthongs in Southern English," Le Maître Phonétique **32**, 1–12.

Juvonen, P., and Koptjevskaja-Tamm, M. (**2016**). *The Lexical Typology of Semantic Shifts* (De Gruyter, London).

Kamiloğlu, R., Fischer, A. H., and Sauter, D. (**2020**). "Good vibrations: A review of vocal expressions of positive emotions," Psychon. Bull. Rev. **27**, 237–265.

Koutseff, A., Reby, D., Martin, O., Levrero, F., Patural, H., and Mathevon, N. (**2018**). "The acoustic space of pain: Cries as indicators of distress recovering dynamics in pre-verbal infants," Bioacoustics **27**(4), 313–325.

Laukka, P., and Elfenbein, H. A. (**2021**). "Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis," Emotion Rev. **13**(1), 3–11.

Levenshtein, V. I. (**1966**). "Binary codes capable of correcting deletions, insertions, and reversals," Sov. Phys. Dokl. **10**(8), 707–710.

List, J.-M., Forkel, R., Greenhill, S. J., Rzymski, C., Englisch, J., and Gray, R. D. (**2022**). "Lexibank, a public repository of standardized wordlists with computed phonological and lexical features," Sci. Data **9**(1), 316.

Morton, E. S. (**1977**). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," Am. Nat. **111**(981), 855–869.

Ohala, J. J. (**1984**). "An ethological perspective on common cross-language utilization of F0 of voice," Phonetica **41**, 1–16.

Ohala, J. J. (**1995**). "The frequency code underlies the sound-symbolic use of voice pitch," in *Sound Symbolism*, edited by L. Hinton, J. Nichols, and J. J. Ohala (Cambridge University Press, Cambridge, UK), pp. 325–347.

Peirce, C. S. (**1955**). "Logic as semiotics: The theory of signs," in *Philosophical Writings of Peirce*, edited by J. Buchler (Dover Publications, Englewood Cliffs, NJ), pp. 98–119.

Perlman, M., Dale, R., and Lupyan, G. (**2015**). "Iconicity can ground the creation of vocal symbols," R. Soc. Open Sci. **2**(8), 150152.

Pisanski, K., Bryant, G. A., Cornec, C., Anikin, A., and Reby, D. (**2022**). "Form follows function in human nonverbal vocalisations," Ethol. Ecol. Evol. **34**(3), 303–321.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C., and Feinberg, D. R. (**2014**). "Vocal indicators of body size in men and women: A meta-analysis," Anim. Behav. **95**, 89–99.

Pisanski, K., Reby, D., and Oleszkiewicz, A. (**2024**). "Humans need auditory experience to produce typical volitional nonverbal vocalizations," Commun. Psychol. **2**(1), 65.

Ponsonnet, M. (**2023**). "Interjections," in *The Oxford Guide to Australian Languages*, edited by C. Bowern (Oxford University Press, Oxford, UK), pp. 564–572.

Ponsonnet, M., and Laginha, K. (**2020**). "The role of the body in descriptions of emotions. A typology of the Australian continent," Pragm. Cogn. **27**(1), 20–83.

Raine, J., Pisanski, K., Simner, J., and Reby, D. (**2019**). "Vocal communication of simulated pain," Bioacoustics **28**(5), 404–426.

Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**(3), 519–530.

Reilly, J., Biun, D., Cowles, W., and Peelle, J. (**2008**). "Where did words come from? A linking theory of sound symbolism and natural language evolution," Nat. Prec. 1–17.

Reiter, S. (**2012**). "Ideophones in awetí," Ph.D. thesis, Köln University, Köln, Germany.

Sapir, E. (**1929**). "A study in phonetic symbolism," J. Exp. Psychol. **12**(3), 225–239.

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (**2010**). "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations," Proc. Natl. Acad. Sci. U.S.A. **107**(6), 2408–2412.

Scherer, K. R. (**2019**). "Acoustic patterning of emotion vocalizations," in *The Oxford Handbook of Voice Perception*, edited by S. Frühholz and P. Belin (Oxford Academic Publishing, Oxford, UK).

Scherer, K. R. (**2021**). "Comment: Advances in studying the vocal expression of emotion: Current contributions and further options," Emotion Rev. **13**(1), 57–59.

Schröder, M. (**2003**). "Experimental study of affect bursts," Speech Commun. **40**(1-2), 99–116.

Schwartz, J. W., and Gouzoules, H. (**2019**). "Decoding human screams: Perception of emotional arousal from pitch and duration," Behavior **156**(13–14), 1283–1307.

Segerer, G., and Flavier, S. (**2023**). "RefLex, reference lexicon of Africa, version 2.2 [dataset]," https://reflex.cnrs.fr/ (Last viewed October 16, 2024).

Spector, F., and Maurer, D. (**2013**). "Early sound symbolism for vowel sounds," I-Perception **4**(4), 239–241.

Valente, D., Magnard, C., Koutseff, A., Patural, H., Chauleur, C., Reby, D., and Pisanski, K. (**2025**). "Vocal communication and perception of pain in childbirth vocalisations," Philos. Trans. R. Soc. B (in press).

Vallery, R., and Lemmens, M. (**2021**). "The sound of taboo," Pragm. Cogn. **28**(1), 87–137.

Vanhove, M. (**2008**). *From Polysemy to Semantic Change* (John Benjamins, Amsterdam).

Voeltz, E. F. K., and Kilian-Hatz, C. (**2001**). *Ideophones: Papers Presented at the 1st International Symposium on Ideophones* (John Benjamins, Amsterdam).

Voronov, M., and Forkel, R. (**2019**). "OneAdder/lingtypology: Release 0.8.5 (v0.8.5) [computer software]," https://doi.org/10.5281/ZENODO.3247007 https://github.com/OneAdder/lingtypology (Last viewed December 14, 2023).

Wichmann, S., Holman, E. W., and Brown, C. H. (**2022**). "The ASJP Database (version 20) [dataset]," https://asjp.clld.org/ (Last viewed March 1, 2024).

Wierzbicka, A. (**1999**). *Emotions across Languages and Cultures: Diversity and Universals* (Cambridge University Press, Cambridge, UK).

Wilkins, D. P. (**1992**). "Interjections as deictics," J. Pragmatics **18**(2), 119–158.

Zalizniak, A. (**2020**). "Database of semantic shifts," https://Datsemshift.Ru/ (Last viewed October 16, 2024).

16 November 2024 15:11:14