Emotional sound symbolism: Languages rapidly signal valence via phonemes

James S. Adelman[1], Zachary Estes[2*], & Martina Cossu[2]

[1] University of Warwick
[2] Bocconi University
*Corresponding author. Address: Bocconi University, Department of Marketing, Via Röntgen 1, Milan 20136, Italy. E-mail: estes@unibocconi.it. Telephone: +39 3466 058 252.

Abstract: Rapidly communicating the emotional valence of stimuli (i.e., negativity or positivity) is vital for averting dangers and acquiring rewards. We therefore hypothesized that human languages signal emotions via individual phonemes (*emotional sound symbolism*), and more specifically that the phonemes at the beginning of the word signal its valence, as this would maximize the receiver's time to respond adaptively. Analyzing approximately 37,000 words across five different languages (English, Spanish, Dutch, German, and Polish), we found emotional sound symbolism in all five languages, and within each language the first phoneme of a word predicted its valence better than subsequent phonemes. Moreover, given that averting danger is more urgent than acquiring rewards, we further hypothesized and demonstrated that phonemes that are uttered most rapidly tend to convey negativity rather than positivity. Thus, emotional sound symbolism is an adaptation providing an early warning system in human languages, analogous to other species' alarm calls.

Keywords: automatic vigilance, emotion, evolution, language, phonology, sound symbolism.

Manuscript Information: 6085 words (main text); 57 references; 1 table; 4 figures; 6 supplementary tables; 1 supplementary text; 1 supplementary data processing and analysis script.

A century ago Saussure declared that “the sign is arbitrary” (Saussure, 1916/2011), arguing that there is no inherent relation between the sound of a word and its meaning. However, subsequent studies have shown that the sounds of words are indeed systematically related to word meaning (Blasi et al., 2016). This sound symbolism was perhaps best illustrated by Köhler (1929): When shown a rounded object and an angular object, and asked which is “takete” and which is “baluma”, the vast majority of people agree that the angular object should be called “takete”. In fact, phonemes systematically convey a range of physical properties such as size and shape (Blasi et al., 2016; Köhler, 1929; Sapir, 1929) and more general syntactic categories such as nouns and verbs (Farmer, Christiansen, & Monaghan, 2006).

Sound symbolism is fundamental to language. It supported language evolution (i.e., the emergence of language), it influences language development (i.e., the emergence and persistence of words within languages), and it facilitates language learning (i.e., learning the words of a language). Regarding language evolution, humans appear biologically predisposed for sound symbolism: Chimpanzees exhibit behavioral precursors of it (Ludwig, Adachi, & Matsuzawa, 2011), preverbal infants and aphasic adults are sensitive to it (Asano et al., 2015; Meteyard et al. 2015), and it is observed across many languages (Blasi et al., 2016; Nuckolls, 1999; Perniss et al., 2010). Sound symbolism may have emerged as a physical analogy between the production of the speech sound and the meaning of the word (Imai & Kita, 2014). For instance, larger animals tend to produce less dispersed, lower pitch vocalizations (Lloyd, 2005), and by analogy, languages tend to use lower-frequency sounds to name larger objects (Sapir, 1929). Regarding language development, sound symbolism is more evident among languages with small vocabularies than with larger vocabularies, and among words learned during childhood than during adulthood (Monaghan, Shillcock, Christiansen, & Kirby, 2014). Sound symbolism also persists within and across modern languages via subtle statistical associations (Blasi et al., 2016;

Farmer et al., 2006; Monaghan et al. 2014). For instance, across thousands of languages the word denoting “dog” is relatively likely to have the /s/ phoneme and unlikely to have /t/ (Blasi et al. 2016). Regarding language learning, sound symbolism facilitates word learning (Imai, Kita, Nagumo, & Okada, 2008) and categorization (Monaghan, Christiansen, & Fitneva, 2011) by linguistically grouping words whose referents are similar. For instance, back vowels like /ɒ/ in “dog” and “hog” group together relatively large objects, whereas front vowels like /æ/ in “cat” and “ant” group together small objects (Sapir, 1929).

Current explanations attribute sound symbolism to pre-existent cognitive processes (for review see Sidhu & Pexman, 2017). For instance, sound symbolism may have emerged from the older cognitive capacities for perceiving analogy (e.g., between articulatory gesture and word meaning) or for detecting statistical associations (e.g., between sounds and objects or actions), as described above. That is, according to this *spandrel account*, sound symbolism is a happy consequence of prior associative and analogical skills; it is a spandrel of those previously evolved cognitive capacities. Humans (or their hominid ancestors) first developed associative and analogical skills, which provided adaptive benefits unrelated to communication. As a side-effect, communicative signals that were sound-symbolic were easier to learn, so communication systems became more sound-symbolic and could contain more distinct signals. This enabled the development of more complex communication systems that became human languages (Imai & Kita, 2014). Thus, sound symbolism occurred because words that obeyed sound symbolic regularities had a survival advantage over words that did not; and not because humans that learned or employed sound symbolic relationships well had a survival advantage over those who learned or employed them poorly.

Could sound symbolism instead have come about precisely because it provided an immediate advantage to humans that used it? We propose an alternative *adaptation account*, in which sound symbolic communication of one especially important property of stimuli – emotion – was selected for due to its adaptive value. Animals communicate about many types of stimuli in their environment (Pollick & De Waal, 2007), but most fundamentally, animals communicate about dangers (e.g., predators, threats) and opportunities (e.g., food, sex), thereby supporting the fitness and survival of individuals and the species (Seyfarth & Cheney, 2003). The cognitive faculties underpinning language presumably emerged and developed in humans from this same evolutionary pressure (Pinker, 1995). At some point, only some humans had cognitive (and/or physiological) capacities that favored sound-symbolic communication systems. These humans could communicate with one another about dangers and opportunities more efficiently than other humans. Given the urgency of this emotional information for survival, especially negative danger-related information, this communicative efficiency endowed its users with a survival advantage, providing selective pressure for the perception and production of sound symbolic communication to become a property of the population as a whole. By this account, sound symbolism is an adaptation rather than a spandrel.

We test the adaptation account by investigating properties of contemporary human languages that it implies. First and most simply, given that dangers and opportunities respectively induce negative and positive affective states in the perceiver (Lindquist et al., 2012; Russell, 2003), we hypothesized that individual phonemes are statistically associated with negative and positive emotion (*emotional sound symbolism*). Furthermore, if emotional sound symbolism was important for our species' survival, then it should be observed across languages. Indeed, some preliminary evidence of emotional sound symbolism has been obtained (Heise, 1966; Louwerse & Qu, in press; Thorndike, 1945), but those studies used non-random samples of words and confounded valence with lexical and/or emotional factors (e.g., word frequency, arousal) that more parsimoniously explain the observed effects (see the Discussion for further detail). However, emotional sound symbolism might also be expected on a spandrel account as another form of sound symbolism; emotional valence is a major semantic dimension, and spandrels may be incidentally adaptive.

The adaptive urgency of communicating dangers and opportunities (Chittka, Skorupski, & Raine, 2009; Kousta, Vinson, & Vigliocco, 2009; Trimmer et al., 2008) yields two further predictions that would appear to be markers of special design for survival-linked communication, consistent with the adaptation account. To our knowledge, these have never been tested. The faster an emotional signal is received, the sooner an adaptive behavior can be executed, and hence the more likely the receiver is to reap rewards (e.g., food) and avert catastrophes (e.g., predation). Words could communicate emotions most rapidly, and hence facilitate vital responding, if the emotion-conveying phonemes were those that are perceived first (e.g., the /s/ sound in “snake”). Thus, we hypothesized that the valence of a word is best predicted by phonemes at the beginning of the word. Further, averting danger is more urgent than acquiring rewards, so negative stimuli are prioritized in human behaviors (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Carretié et al., 2004; Fox, Russo, Bowles, & Dutton, 2001; Smith, Cacioppo, Larsen, & Chartrand, 2003). We thus hypothesized that “fast” phonemes (i.e., those that are uttered most quickly) convey negativity, as this would maximize the listener’s time to avert potentially lethal dangers.

We used the five languages for which the largest datasets of emotion ratings for randomly sampled words are currently available for our tests of the adaptation account: English (N = 12,847 words), Spanish (N = 13,935), Dutch (N = 4270), German (N = 2900), and Polish (N = 2902). Within each language we conducted hierarchical regression analyses testing whether the individual phonemes of words (i.e., counts of each phoneme for each word) predict the emotional valence of those words, after statistically controlling important affective (i.e., arousal) and lexical factors (i.e., word length, frequency, and contextual diversity). For comparison, we also tested for emotional sound symbolism at the level of *phonetic features* (i.e., place and manner of articulation for consonants, place and height for vowels, and voicing). In English, for instance, all nasal phonemes (i.e., /m/, /n/, and /ŋ/) were grouped, and other phonetic features were similarly grouped (e.g., bilabials, voiceless consonants, front vowels, etc.). We then tested whether individual phonemes predict word valence better than those general phonetic features.

Methods

Data

Within each language we first retrieved the largest available database of arousal and valence ratings, we then retrieved phonemic transcriptions and lexical control variables (word frequency and contextual diversity, both log-transformed after adding 1 to deal with zero counts), and finally we counted word length (number of letters, number of consonants, number of vowels). For generality across languages, all phonemes are transcribed using the International Phonetic Alphabet (*IPA*).

English. We retrieved arousal and valence ratings for 13,915 English words (Warriner, Kuperman, & Brysbaert, 2013), and we retrieved phonemic transcriptions from the Carnegie Mellon University pronouncing dictionary (*CMU*), hand-coding a further 152 items (mostly closed compounds, e.g., “applejack"), and word frequency and contextual diversity (Subtlex-US; Brysbaert & New, 2009) from the English Lexicon Project (*ELP;* Balota et al., 2007). Excluding 1068 words that did not appear in either our CMU-based transcription list or that are open compounds (e.g., “soda pop”), the final list included 12,847 words and 39 phonemes. We also retrieved pronunciation latencies for 12,595 of those words from ELP. The pronunciation task (a.k.a. naming or reading aloud) entails participants viewing single words on screen and saying them aloud as quickly as possible, and pronunciation latencies are the time from word onset to voice onset (i.e., the initial sound of the pronunciation). Finally, an independent replication study included a set of 2820 monosyllabic words (Adelman, Marquis, Sabatos-DeVito, & Estes, 2013), which is the next-largest set of affective ratings of English words. Excluding 47 words that did not appear in either our CMU-based transcription list, this analysis included 2773 words.

Spanish. We retrieved arousal and valence ratings for 14,031 Spanish words (Stadthagen-Gonzalez, Imbault, Sánchez, & Brysbaert, 2016), and we retrieved phonemic transcriptions, word frequency, and contextual diversity from EsPal (Duchon et al., 2013). Excluding 96 words that did not have a phonemic transcription in EsPal, the final list included 13,935 words and 31 phonemes.

Dutch. We retrieved arousal and valence ratings for 4299 Dutch words (Moors et al., 2013), and we retrieved phonemic transcriptions from CELEX (Baayen, Piepenbrock, & Van Rijn, 1993) and word frequency and contextual diversity from Subtlex-NL (Keuleers, Brysbaert, & New, 2010). Excluding 29 words that did not appear in either CELEX, the final list included 4270 words and 42 phonemes.

German. We retrieved arousal and valence ratings for 2902 German words (Vo et al., 2009), and we retrieved phonemic transcriptions from CELEX (Baayen et al., 1993) and word frequency from Subtlex-DE (Brysbaert et al., 2011). No measure of contextual diversity was available in German. Excluding 2 words that did not appear in CELEX, the final list included 2900 words and 47 phonemes. We also retrieved pronunciation latencies for 648 of those words from Schröter and Schroeder (2017). Note that because Schröter and Schroeder investigated reading among both children and adults, their sampled words were simple and highly frequent (so that children could read them easily), but our analyses used only the pronunciation latencies by young adults (not children).

Polish. We retrieved arousal and valence ratings for 2902 Polish words (Riegel et al., 2015), then we constructed phonemic transcriptions following the transparent spelling-sound rules of Polish, so these were available for all words, and we retrieved word frequency and contextual diversity from Subtlex-PL (Mandera, Keuleers, Wodniecka, & Brysbaert, 2014). 36 phonemes were available for analysis in the database.

Analyses

Within each language we first counted the number of each phoneme that was present in each given word. In English for instance, “dog” had scores of 1 for the /d/, /ɒ/, and /g/ phonemes, and had scores of 0 for the remaining 36 phonemes. This produced a 39 (phonemes) × 12,847 (words) matrix coding all phonemes of all words in our English dataset. In addition, we coded each phoneme’s position within the word (e.g., first or last phoneme) and its typical phonetic features (i.e., place and manner of articulation for consonants; place and height for vowels; voicing), and we repeated this procedure for each of the five languages.

We then conducted a series of hierarchical linear regressions separately within each language, with valence ratings as the dependent variable in all analyses unless otherwise stated. In all cases we entered word length (number of letters, number of consonants, and number of vowels), log. frequency, log. contextual diversity (except in German), and arousal in a first block, and then we entered the phonemes in a second block. This allowed us to test whether the phonemes explained a significant amount of unique variance in valence ratings after statically accounting for the control variables (i.e., length, frequency, diversity, and arousal). For instance, in English the second block of the main regression included all 39 phonemes as predictors of valence, thus revealing which phonemes are significantly associated with positive valence (positive coefficient in a weighted deviation contrast) or negative valence (negative coefficient).

All variables were centered, and hence the phoneme coefficients indicate the difference from the average phoneme.[1]

Within each language we conducted the same four analyses. (i) *All phonemes*: The phoneme variables were counts of phoneme occurrence regardless of position within the word. (ii) *Phonetic features*: The phonetic feature variables were counts of inferred phonetic feature occurrences regardless of position within the word. (iii) *First phoneme*: The phoneme variables dummy coded whether the each given phoneme was the initial phoneme of the given word (0 = no, 1 = yes). (iv) *Last phoneme*: The phoneme variables dummy coded whether the each given phoneme was the final phoneme of the given word (0 = no, 1 = yes). Note that some phonemes never occurred in the first or the last position. In English for instance, the /ŋ/ phoneme (as in "thi**ng**") never occurs as the first phoneme of a word. The *first phoneme* and *last phoneme* analyses therefore include slightly fewer phonemes than the *all phonemes* analyses, indicated by empty cells in Supplementary Tables 1-6.

Linguistic analyses such as this typically assume the standard α of .05 for identifying significantly predictive phonemes (e.g., Heise, 1966; Louwerse & Qu, in press; Thorndike, 1945). We also adopt this standard, because the error rate for the omnibus *F*-test (i.e., the R^2 change when adding the phoneme block to the control block) is fixed at .05, and we would not interpret the individual coefficients without a significant omnibus effect. So to be clear, the percentage of significant phonemes in each analysis reported below should be compared to the

[1] For analyses of phonemes at a particular position (e.g., first phoneme), centering the variables resulted in "simple coding", so the phoneme coefficients indicate the difference from the average phoneme. A different type of contrast was needed to obtain a similar interpretation for analyses with phoneme counts across the whole word that both (a) were not confounded with a general effect of number of phonemes and (b) provided an independent test of each phoneme differing from the average phoneme. For each phoneme of interest, the null hypothesis that the phoneme of interest did not differ from the average phoneme was instantiated by, for each word, redistributing the count for the phoneme of interest across the other phonemes in proportion to their frequency in the language. The alternative was then tested (and coefficient derived) by including the count of the phoneme of interest as an additional variable in the model. This statistical method has no bearing on the results of the overall models reported in the main text; it merely adjusts the individual phonemes' coefficients so that they indicate their difference from the average phoneme.

null-expected 5%. Given that our largest analysis (German language, *all phonemes*) included 47 phonemes, fewer than 2.5 phonemes in each analysis were expected to be significant by random chance alone. Nonetheless, to provide a more statistically conservative view of the data, in all tables we also identify the phonemes that remained significant even after a Bonferroni correction for family-wise error rate.

Results

Full results for each language, including the lexical and affective control factors entered in the first block of the regression analyses, are reported in Supplementary Tables 1-6. Here we describe our results of interest, namely, the second block of the regression analyses in which phonemes were included as predictors of word valence. Effect sizes are reported as regression coefficients indicating the relative valence of each phoneme, and R^2 for the phonemes' collective effect.

Emotional sound symbolism. In English, phonemes collectively explained a significant amount of unique variance in valence ratings (R^2 = 1.44%, $P < .001$), and 36% of individual phonemes were significantly negative or positive (Fig. 1), thus demonstrating emotional sound symbolism. For comparison, after statistically controlling all other factors, word length explained substantially less variance in valence ratings (R^2 = 0.06%), thus revealing that the effect of phonemes on valence was relatively large. We tested the reliability and robustness of this effect in four ways. First, to test inter-item reliability, 1000 times we randomly split the word pool into halves, replicating the *all phonemes* analysis on both halves and then testing the correlation of the phonemes' regression coefficients for each of those 1000 split-samples. This inter-item reliability was good (average r = .52, reliability = .68). Second, to test whether the result was due to common affixes (e.g., dis-, -est) we again replicated the *all phonemes* analysis, but including only monomorphemic words (N = 5710), which lack affixes. The result replicated (R^2 = 1.19%, *P* < .001), confirming that emotional sound symbolism was not attributable to prefixes (e.g., dis-) or suffixes (-est). Third, we also replicated the result with an independent dataset of emotion ratings for 2773 words (Adelman et al., 2013; R^2 = 2.26%, *P* < .001). Fourth, to test inter-rater reliability, we included only the 2184 words that occurred in both the original dataset (Warriner et al., 2013) and the replication dataset (Adelman et al., 2013). We replicated the *all phonemes* analysis separately within both datasets, and we then tested the correlation between the phonemes' coefficients from these two independent analyses of the same words rated by different groups of participants. This inter-rater reliability was good (*r* = .86, *P* < .001). Thus, in English, emotional sound symbolism was highly robust and reliable across participants, items, and datasets.

Phonemes also significantly predicted valence ratings in Spanish (R^2 = 1.40%, *P* < .001), Dutch (R^2 = 2.29%, *P* < .001), German (R^2 = 2.81%, *P* < .001), and Polish (R^2 = 4.32%, *P* < .001), and relatively high percentages of individual phonemes significantly predicted valence in Spanish (45%; Fig. 1), Dutch (21%), German (21%), and Polish (33%). These percentages are markedly higher than the 5% of phonemes that would be expected to reach significance by chance alone. See Supplementary Tables 1-6.

We also tested whether this emotional sound symbolism could be explained more parsimoniously by phonetic features such as place of articulation (e.g., linguodentals), manner of articulation (e.g., fricatives), vowel height (e.g., low vowels), voicing, and so on. Phonetic features did significantly predict word valence in English (R^2 = 0.63%, *P* < .001), Spanish (R^2 = 0.35%, *P* < .001), Dutch (R^2 = 0.86%, *P* = .004), German (R^2 = 1.65%, *P* < .001), and Polish (R^2 = 1.79%, *P* < .001). Within each language, however, those effects of phonetics features were substantially smaller than the effects of individual phonemes reported above (see Fig. 2). Moreover, even after statistically accounting for those phonetic features (entered as a second block in the regression), the individual phonemes (entered as a third block) still significantly predicted word valence within each language: English (R^2 = 0.81%, P < .001), Spanish (R^2 = 1.05%, P < .001), Dutch (R^2 = 1.43%, P < .001), German (R^2 = 1.15%, P = .013), and Polish (R^2 = 2.54%, P < .001). In English, for example, /f/ is significantly positive whereas /s/ is significantly negative (see Supplementary Table 1). Simply treating them both as fricatives fails to capture this critical difference. Thus, emotional sound symbolism occurs at the level of individual phonemes rather than general phonetic features.

Front-loading. To examine the temporal dynamics of emotional sound symbolism within words, we also predicted valence ratings from the phonemes in a particular position within the word. First phoneme significantly predicted valence in each of the five languages: English (R^2 = 2.14%, P < .001), Spanish (R^2 = 1.16%, P < .001), Dutch (R^2 = 2.12%, P < .001), German (R^2 = 2.89%, P < .001), and Polish (R^2 = 3.86%, P < .001). However, last phoneme significantly predicted valence only in English (R^2 = 0.48%, P < .001), Spanish (R^2 = 0.77%, P < .001), Dutch (R^2 = 0.85%, P = .041), and German (R^2 = 1.75%, P < .001), and first phoneme predicted valence better than last phoneme within each of the five languages (Fig. 3A). See Table 1 for examples in English. To examine this phoneme position effect in greater detail, we also calculated the independent contributions of each of the first five phonemes of words. That is, we calculated the ΔR^2 when only the first phoneme was added to the control block, when only the second phoneme was added, and so on. Only the English (N = 9523 words) and Spanish (N = 12,917) datasets included sufficient numbers of words with at least five phonemes to support this analysis. Each phoneme position was analyzed independently, not cumulatively, so the residual variance available to-be-explained was constant across phoneme positions. Yet, the contribution of individual phonemes to the words’ valence decreased with the phonemes’ distance from the beginning of the word (Fig. 3B). First phonemes predicted valence better than second phonemes, and so on.

The preceding analyses demonstrate that emotional sound symbolism is strongest at the beginnings of words, but they do not indicate whether this front-loading is distinctive to valence. Might front-loading be a general principle of the relation between sound and meaning? Preliminary evidence suggests not: Monaghan and Christiansen (2006) showed that the grammatical category of a word (i.e., nouns versus verbs) is better predicted by its final phonemes than by its initial phonemes. Nonetheless, we conducted another test of this hypothesis using perhaps the best known and most influential lexico-semantic property: word frequency. We replicated the preceding analyses, but with (log) word frequency as the dependent variable instead of word valence, and without contextual diversity as a control factor (due to its strong collinearity with frequency; Adelman, Brown, & Quesada, 2006). These tests were conducted in English and Spanish, the two largest datasets. After accounting for the lexical and affective control factors (length and arousal), the first phoneme of a word did significantly predict its frequency within both English (R^2 = 0.64%, $P < .001$) and Spanish (R^2 = 0.70, $P < .001$). Crucially however, word frequency was predicted as well (or better) by the last phoneme of the word in both English (R^2 = 0.85%, $P < .001$) and Spanish (R^2 = 0.96%, $P < .001$). Thus, front-loading does not appear to be a general principle of sound symbolism; it appears to be a relatively distinctive property of emotional sound symbolism.

Negative priority. Finally, to examine temporal differences between negative and positive phonemes, we calculated the average pronunciation latency of each phoneme (i.e., the time from stimulus onset to voice onset), and used those pronunciation latencies to predict the average valence of each phoneme. Currently, English and German are the only languages for which datasets of pronunciation latencies (Balota et al., 2007; Schröter & Schroeder, 2017) contain enough words that also have valence and arousal ratings available to support this analysis. We used each phoneme's coefficient from the *first phoneme* analysis as a measure of its relative valence, because pronunciation latencies measure voice onset (i.e., time until first phoneme). Then we similarly determined each phoneme's covariate-adjusted pronunciation latency by replicating that *first phoneme* analysis, but with pronunciation latencies as the dependent variable and valence ratings as an additional control variable (English: R^2 = 48.00%, $P < .001$; German: R^2 = 63.39%, $P < .001$). Pronunciation latencies were shorter in German than in English due to more consistent spelling-sound correspondences in German than in English, and to the use of simpler words in the German dataset (see "Data" above). We used each phoneme's coefficient from this analysis as a measure of its relative pronunciation latency. In English, individual phonemes' pronunciation latency significantly predicted their valence, $N = 36$, $r = .55$, $P < .001$, or $r = .34$, $P = .048$, when one outlier (/ʒ/) is excluded.[2] Pronunciation latency also significantly predicted valence in German, $N = 34$, $r = .42$, $P = .013$. Phonemes that are uttered quickly tend to occur at the beginning of negative words, whereas phonemes that are uttered slowly typically begin positive words (Fig. 4).

Discussion

Prior research has demonstrated sound symbolism for physical properties of objects and actions (e.g., size and shape; Asano et al., 2015; Blasi et al., 2016; Imai & Kita, 2014; Imai et al., 2008; Köhler, 1929; Monaghan et al., 2011; Monaghan et al., 2014; Nuckolls, 1999; Sapir,

---

[2] The negative priority hypothesis suggests that the association of first phonemes to valence is driven by the time taken to communicate these phonemes, that is, from the speaker seeing the stimulus to the hearer identifying the phoneme. Pronunciation latencies (used in the analysis reported above) are an imperfect approximation of communication times because they represent phoneme onsets, whereas the perceiver needs to hear the phoneme – not just its onset – in order to identify it. An alternative approximation therefore is the time from stimulus onset to phoneme offset (i.e., pronunciation latency plus phoneme duration). Unfortunately, such offset latencies are currently available only for 21 English consonants (Rastle, Croot, Harrington, & Coltheart, 2005). Although those offset latencies predicted the corresponding valence coefficients ($r = .32$) similarly to the analysis with pronunciation latencies, this analysis lacked the statistical power ($N = 21$) to reach significance ($P = .15$).

1929). The present research, in contrast, demonstrates sound symbolism for positive and negative emotional states. Analyzing approximately 37,000 randomly sampled words across five languages, we provide strong evidence of emotional sound symbolism that was highly reliable across words, individuals, datasets, and languages. Emotional sound symbolism was also highly specific, occurring at the level of individual phonemes rather than general phonetic features.

The observations of emotional sound symbolism in three Germanic languages (English, Dutch, and German), a Romance language (Spanish), and a Balto-Slavic language (Polish) suggest that emotional sound symbolism may be a general mechanism of the language faculty. Indeed, given that sound symbolism is generally weaker among such Indo-European languages than among most African, Asian, and South American languages (Perniss et al. 2010), we speculate that the emotional sound symbolism demonstrated here in Indo-European languages may be stronger in other language families.

We showed for the first time that emotional sound symbolism is front-loaded: (i) the very first phoneme significantly predicts the valence of the word, (ii) the farther a phoneme is from the beginning of the word, the less well it predicts its valence, and (iii) the first phoneme predicts valence better than the last phoneme within each of the five languages. Effective communication of dangers and opportunities is vital for the fitness and survival of a species (Chittka et al., 2009; Seyfarth & Cheney, 2003; Trimmer et al., 2008). This front-loading of the emotional signal appears to be a particularly efficient means of such emotional communication, especially maximizing the receivers' time to avert dangers by communicating negative information. Notably, front-loading appears to be a relatively unique property of emotional sound symbolism, as Monaghan and Christiansen (2006) demonstrated that grammatical category is back-loaded, and we further showed that word frequency is not front-loaded either.

Finally, we also showed for the first time that emotional sound symbolism prioritizes negative valence: Phonemes that are pronounced most rapidly tend to occur at the beginning of negative words, whereas phonemes pronounced slowly tend to begin positive words.[3] This negative priority in sound symbolism, while novel, is consistent with a broader principle of behavioral adaptation that avoiding negative outcomes is more urgent than obtaining positive outcomes (Baumeister et al., 2001): Evolutionary arms races are asymmetrical between predators gaining food and prey avoiding death (Dawkins, 1982); losses loom larger than gains in risky choice (Kahneman & Tversky, 1979); and negative stimuli capture attention earlier and hold attention longer than other stimuli (Carretie et al., 2004; Fox et al., 2001). This preferential attention to negative stimuli, known as *automatic vigilance*, can have divergent effects on word processing and judgment. In tasks akin to alarm signaling, such as judging whether a word is negative or positive, negative words evoke faster responding than positive words. However, in tasks for which valence is irrelevant, such as lexical decisions, negative words instead evoke slower responding (Estes & Verges, 2008). The reading aloud task is particularly interesting in this regard. On one hand, reading aloud resembles an alarm calling task. But on the other hand, reading aloud studies are nowadays typically conducted in such a way to prevent alarm calling: The participant is alone at a computer, in a sterile laboratory setting, with nobody listening. In this context, negative words evoke slower responding than positive words (Estes & Adelman, 2008-a, 2008-b; Kuperman et al., 2014). Thus collectively the past and present studies suggest that negative phonemes are uttered more rapidly, but negative words may only be pronounced more rapidly in contexts that evoke alarm signaling. When the context prevents alarm signaling,

[3] One could argue that it is not valence *per se* that matters for adaptive responding, but rather an interaction of valence and arousal (e.g., Robinson et al., 2004), such that highly arousing negative stimuli (e.g., “snake”) are behaviorally prioritized but low-arousal negative stimuli (e.g., “coffin”) are not. We additionally tested such an interaction model of emotional sound symbolism, but because the evidence for this model was inconclusive, full details of the analysis are reported as Supplementary Materials.

the negative valence distracts the responder's attention from the task at hand, thus slowing responses to negative words. Though speculative, this account shows that negative advantages and disadvantages could be understood within a single framework (whose details are yet to be confirmed).

Recently Louwerse and Qu (in press) found that nasal phonemes as the first consonant (e.g., "unable") predict negative valence in English, German, and Dutch but positive valence in Chinese. Our study provides important theoretical advances beyond Louwerse and Qu's study. Most importantly, our study provides greater explanatory power. Because their analyses were exploratory, they lacked theoretical predictions or explanations of why or how phonetic features might predict valence. In contrast, we developed a theory of sound symbolism based on language evolution and adaptive behavior, which led us not only to predict emotional sound symbolism, but also to generate two novel hypotheses about the front-loading and negative priority of emotional phonemes. A second critical difference is that our study has both greater specificity and greater generality. Louwerse and Qu examined general phonetic features, whereas we showed that emotional sound symbolism is better explained by specific phonemes. At the same time, our analyses are also more general: Whereas they identify a single phonological feature (i.e., nasals as the first consonant) that predicts valence, we identify dozens of phonemes that predict valence.

Our study also provides fundamental methodological improvements. For instance, Louwerse and Qu (in press) analyzed non-random samples of words. Using an "extreme samples" approach, they excluded the 60% of words in the middle of the valence range, thereby substantially reducing the sample size and limiting the generalizability of the results. We instead analyzed, in their entirety, the largest available samples of randomly selected words. Moreover, Louwerse and Qu did not control for word length, frequency, or arousal, all of which correlate with valence (e.g., Kuperman et al., 2014), and any of which thus may account for their observed effect. In contrast, our analyses first accounted for these control factors, and only then did we test whether phonemes predicted any residual variance in valence ratings. Louwerse and Qu also did not account for morphological redundancy in their analyses (e.g., "unable", "unfit", "unkind", "unwell" etc.), and indeed this appears to explain their effect.[4] In contrast, we included additional analyses of monomorphemic words only, thereby eliminating morphological redundancy, and our results replicated. In sum, the present research contributes a well-specified theoretical model with novel predictions and high explanatory power, and uses larger and more representative samples with better controlled and more robust analyses to provide more specific, more general, and more generalizable insights about emotional sound symbolism.

The prevailing account explains sound symbolism as a side-effect of other, pre-existing cognitive faculties (Imai & Kita, 2014; Monaghan et al., 2014; Perniss & Vigliocco, 2014) such as analogy and generalization (see Sidhu & Pexman, 2017). Particular sounds become associated with specific physical properties by analogy (e.g., large objects make low-pitched sounds) and/or generalization (e.g., small objects tend to be named with the /æ/ sound). Humans possessed these general cognitive faculties (i.e., analogy and generalization), which then supported the emergence of sound symbolism and eventually language.

[4] Further analyses show that Louwerse and Qu's finding that initial nasal consonants are negative in English, Dutch, and German is attributable to negating prefixes such as "in-" and "un-" in English and German and "on-" in Dutch. To begin with, our analyses show that nasal phonemes in the first position of the word (e.g., "nose") do *not* predict valence in English (Table 1), Dutch (Supplementary Table S4), or German (Supplementary Table S5). Rather, /n/ in the *second* position within the word predicts valence, but /m/ does not. Louwerse and Qu's finding thus arises from the fact that negating prefixes in Germanic languages frequently contain /n/ in the second position, such as "inedible" and "unable". Louwerse and Qu additionally report three behavioral experiments in which native Dutch speakers evaluated words with a nasal or a non-nasal phoneme in the first position. Unfortunately in their Experiments 1 and 2 the response key to indicate negativity (the M key) itself denotes a nasal phoneme. Thus these experiments show that Dutch speakers are more likely to evaluate words with initial nasals (e.g., "meat") by pressing a nasal key (M) than a non-nasal key (Z key). In their Experiment 3 Dutch speakers showed no effect of initial nasals on valence judgments. Thus Louwerse and Qu provide no clear evidence that initial nasals are associated with negativity in Germanic languages (see their paper for evidence of a positive association in Chinese).

We propose a more radical account that instead explains sound symbolism as an *adaptation*. We argue that sound symbolism for affective information directly improved our species' adaptive fitness by providing two important advantages for communicators: By phonologically contrasting dangers and opportunities, emotional sound symbolism (i) reduces potentially fatal miscommunications and (ii) facilitates rapid responding to vital stimuli, especially dangers. Given our evidence that certain phonemes signal potential danger and others signal opportunity, the receiver is less likely to confuse dangers with opportunities. Moreover, because the very first phoneme predicts the valence of the word, the receiver can prepare and initiate an adaptive response even before the word and its referent are fully identified. Furthermore, because negative phonemes are pronounced more rapidly than positive phonemes, the receiver gains maximum time to avert potentially fatal dangers, which is more urgent than acquiring rewards (Baumeister et al. 2001; Fox et al., 2001). The spandrel view could simply explain the occurrence of emotional sound symbolism in the same way it explains sound symbolism for size and shape: by exaptation from analogy and/or generalization. Notably, however, the spandrel view fails to predict or explain the temporal characteristics of emotional sound symbolism that we demonstrated.

Our view that (emotional) sound symbolism is an adaptation implies that some cognitive (or physiological) change occurred in human capacities. This could be as simple as a specific bias to learn sound symbolic structures in communication systems or an improvement in generalization ability that was favored because it produces emotional sound symbolism. However, a more substantial change may have been involved. In particular, the adaptive value of emotional sound symbolism relies on phonemes. Emotional sound symbolism produces an alarm signaling system that has an efficiency advantage via negative priority. This is possible because phonemes are short, and can combine to produce different signals with the same beginning. By contrast, there is limited evidence for phoneme-like combinatorial abilities in non-human species, only with elements that are syllable-like in duration and content (Engesser et al., 2015). As such, we conjecture that the efficiency advantage given by emotional sound symbolism may have been the specific adaptive advantage in communication that produced a selection pressure for humans to perceive and produce phonemes.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814-823.

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129,* 530-535.

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. J., & Estes, Z. (2013). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 39(4), 1037-1053.

Asano, M. et al. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, 63, 196-205.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical data base on CD-ROM*.

Balota, D. A. et al. (2007). The English lexicon project. *Behavioral Research Methods,* 39(3), 445-459.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.

Blasi, D. E., Wichmann, S., Hammarstrom, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818-10823.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods*, 41(4), 977-990.

Brysbaert, M. et al. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58, 412-424.

Carretié, L., Hinojosa, J. A., Martín-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: Neural correlates. *Human Brain Mapping*, 22, 290-299.

Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed–accuracy tradeoffs in animal decision making. *Trends in Ecology and Evolution,* 24(7), 400-407.

Dawkins, R. (1982). *The extended phenotype.* Oxford: Oxford University Press.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavioral Research Methods,* 45(4), 1246-1258.

Engesser, S., Crane, M. S., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental evidence for phoneme structure in the vocal system of a non-human. *PLoS Biology 13(6):* e1002171.

Estes, Z. & Adelman, J. S. (2008-a). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion, 8,* 441-444.

Estes, Z. & Adelman, J. S. (2008-b). Automatic vigilance for negative words is categorical and general. *Emotion, 8,* 453-457.

Estes, Z., & Verges, M. (2008). Freeze or flee? Negative stimuli elicit selective responding. *Cognition, 108*(2), 557-565.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences,* 103(32), 12203-12208.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety?. *Journal of Experimental Psychology: General*, *130*(4), 681-700.

Heise, D. R. (1966). Sound-meaning correlations among 1,000 English words. *Language and Speech*, 9(1), 14-27.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130298.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54-65.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavioral Research Methods*, 42(3), 643-650.

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

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3), 473-481.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065-1081.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, 35(03), 121-143.

Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America*, 117, 944-1005.

Louwerse, M., & Qu, Z. (in press). Estimating valence from the sound of a word: Computational, experimental, and cross-linguistic evidence. *Psychonomic Bulletin & Review*.

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences*, 108(51), 20661-20665.

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2014). Subtlex-PL: subtitle-based word frequency estimates for Polish. *Behavioral Research Methods,* 47(2), 471-483.

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia, 76,* 264-275.

Monaghan, P., & Christiansen, M. H. (2006). Why form-meaning mappings are not entirely arbitrary in language. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1838-1843). Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140(3), 325-347.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 369(1651), 20130299.

Moors, A. et al. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavioral Research Methods,* 45(1), 169-177.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology,* 28, 225-252.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 369(1651), 20130300.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology,* article 227.

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Penguin, UK).

Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, 106(42), 18010-18015.

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1083-1095.

Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., & Marchewka, A. (2015). Nencki Affective Word List (NAWL): the cultural adaptation of the Berlin Affective Word List–Reloaded (BAWL-R) for Polish. *Behavioral Research Methods,* 47(4), 1222-1236.

Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin, 30*, 1472-1484.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145-172.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3), 225-239.

Saussure, F-D (1916/2011) *Course in general linguistics* (Columbia University Press, New York) W. Baskin Trans.

Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods, 49,* 2183-2203.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, 54(1), 145-173.

Sidhu, D. M., & Pexman, P. M. (2017). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, 1-25.

Smith, N. H., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, 41(2), 171-183.

Stadthagen-Gonzalez, H., Imbault, C., Sánchez, M. A. P., & Brysbaert, M. (2016). Norms of valence and arousal for 14,031 Spanish words. *Behavioral Research Methods,* in-press DOI: 10.3758/s13428-015-0700-2.

Thorndike, E. L. (1945). The association of certain sounds with pleasant and unpleasant meanings. *Psychological Review*, 52(3), 143-149.

Trimmer, P. C. et al. (2008). Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 275, 2353-2361.

Vo, M. L. H. et al. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavioral Research Methods,* 41(2), 534-538.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavioral Research Methods*, 45(4), 1191-1207.

Table 1. The first phoneme of a word predicts the word's valence in English, after controlling for lexical and affective factors (i.e., word length, frequency, contextual diversity, and arousal). Example words are representative of the given phoneme's valence (e.g., positive phonemes are exemplified with positive words). Significant predictors (*P* < .05) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

First Phoneme	Example	Valence *B*	Valence *t*	Valence *p*
ɑ:	awe	0.20	1.91	.056
æ	at	0.01	0.19	.850
ʌ	ugly	-0.24	-4.66	***.000***
ɔ:	ought	0.08	0.59	.554
aʊ	ounce	0.07	0.39	.695
aɪ	idea	0.38	2.19	**.028**
b	bag	-0.01	-0.33	.739
t͡ʃ	cheer	0.40	3.62	***.000***
d	die	-0.49	-11.28	***.000***
ɛ	enjoy	0.38	5.07	***.000***
ɝ:	earth	0.54	2.60	**.009**
eɪ	ape	-0.10	-0.48	.630
f	fun	0.14	2.86	**.004**
g	good	0.12	1.95	.052
h	hate	-0.13	-2.21	**.027**
ɪ	ill	-0.21	-4.36	***.000***
i:	emu	0.34	1.47	.143
d͡ʒ	joy	0.31	3.27	***.001***
k	cap	0.06	1.84	.066
l	lid	0.12	1.88	.060
m	mat	-0.02	-0.49	.622
n	net	0.04	0.56	.572
oʊ	odor	-0.41	-3.04	**.002**
ɔɪ	oily	-0.78	-1.29	.198
p	peace	0.12	3.12	**.002**
r	rug	0.03	0.71	.479
s	stem	0.04	1.20	.229
ʃ	shake	-0.18	-1.83	.067
t	tin	-0.04	-0.77	.440
θ	three	0.10	0.70	.485
u:	oodles	0.42	0.59	.552
v	victory	0.21	2.52	**.012**
w	wall	0.00	0.05	.958
j	unique	0.37	2.45	**.014**
z	zoo	0.72	2.56	**.010**
ʒ	genre	1.02	0.84	.401

Fig. 1. Individual phonemes predict word valence. Each phoneme is plotted with a vertical position defined by the regression coefficient predicting word valence (higher phonemes are more positive) and a horizontal position defined by the precision of this coefficient (1/SE; phonemes to the right have more precise estimates of their valence). Phonemes within the dark regions are individually significant predictors of word valence ($p < .05$), and the intermediate-shaded region is $.1 > p > .05$. Fewer than two phonemes within each language were expected to reach significance by chance alone. Results are illustrated for phonemes in English and Spanish, the two largest datasets with the most precise valence estimates.

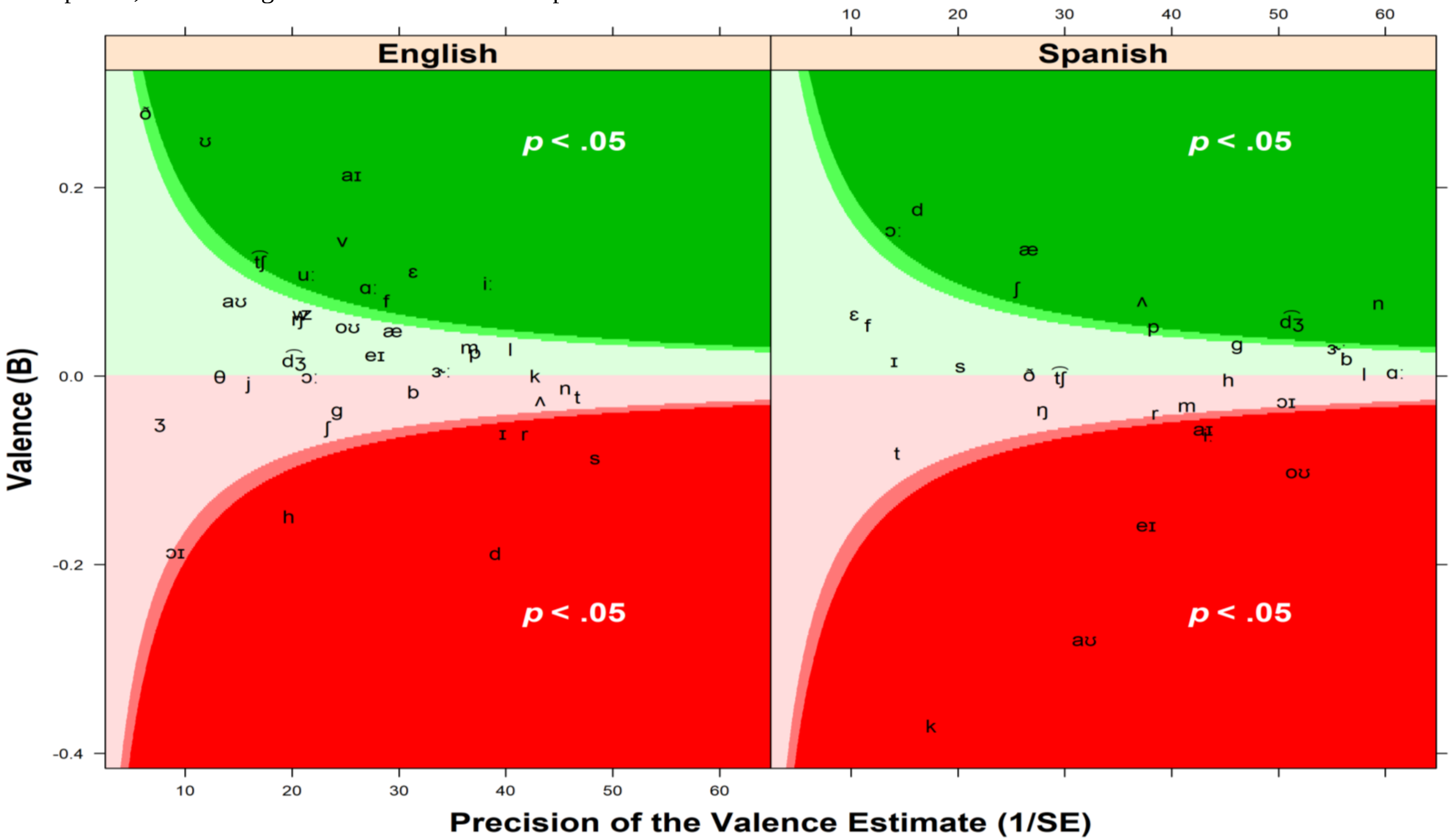


Fig. 2. Individual phonemes predict word valence better than general phonetic features. Effect size (ΔR^2 when phonemes or features are added to control block) of individual phonemes and phonetic features as predictors of the word's valence rating within each language.

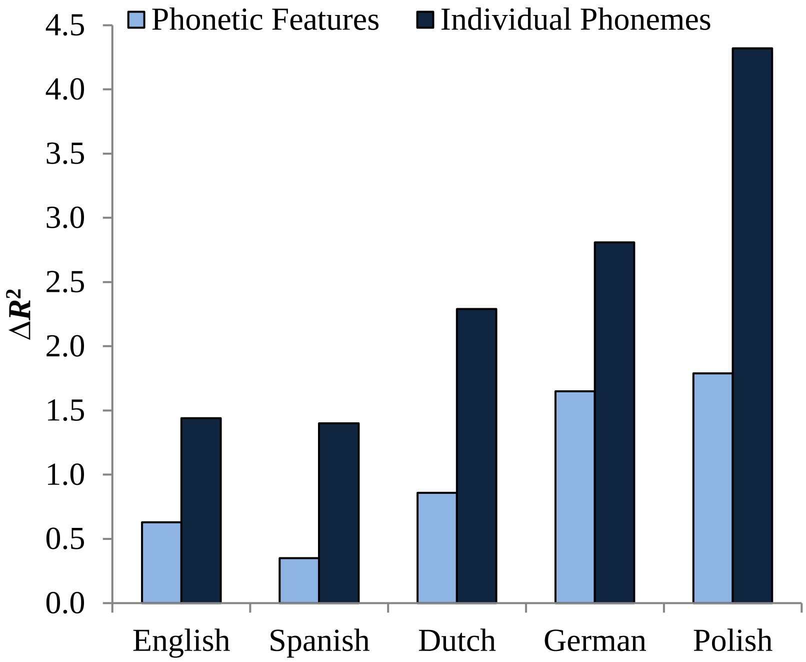


Fig. 3. Phonemes at the beginning of the word predict valence most strongly. (A) Effect size (ΔR^2 when phonemes are added to control block) of the first phoneme and the last phoneme of the word as predictors of the word's valence rating within each language. **(B)** Effect size of the first five phonemes of the word as predictors of the word's valence rating among English (N = 9523) and Spanish (N = 12,917) words with at least five phonemes. Note that each phoneme position was analyzed independently, not cumulatively, so the residual variance available to-be-explained was constant across phoneme positions.

A. **B.**

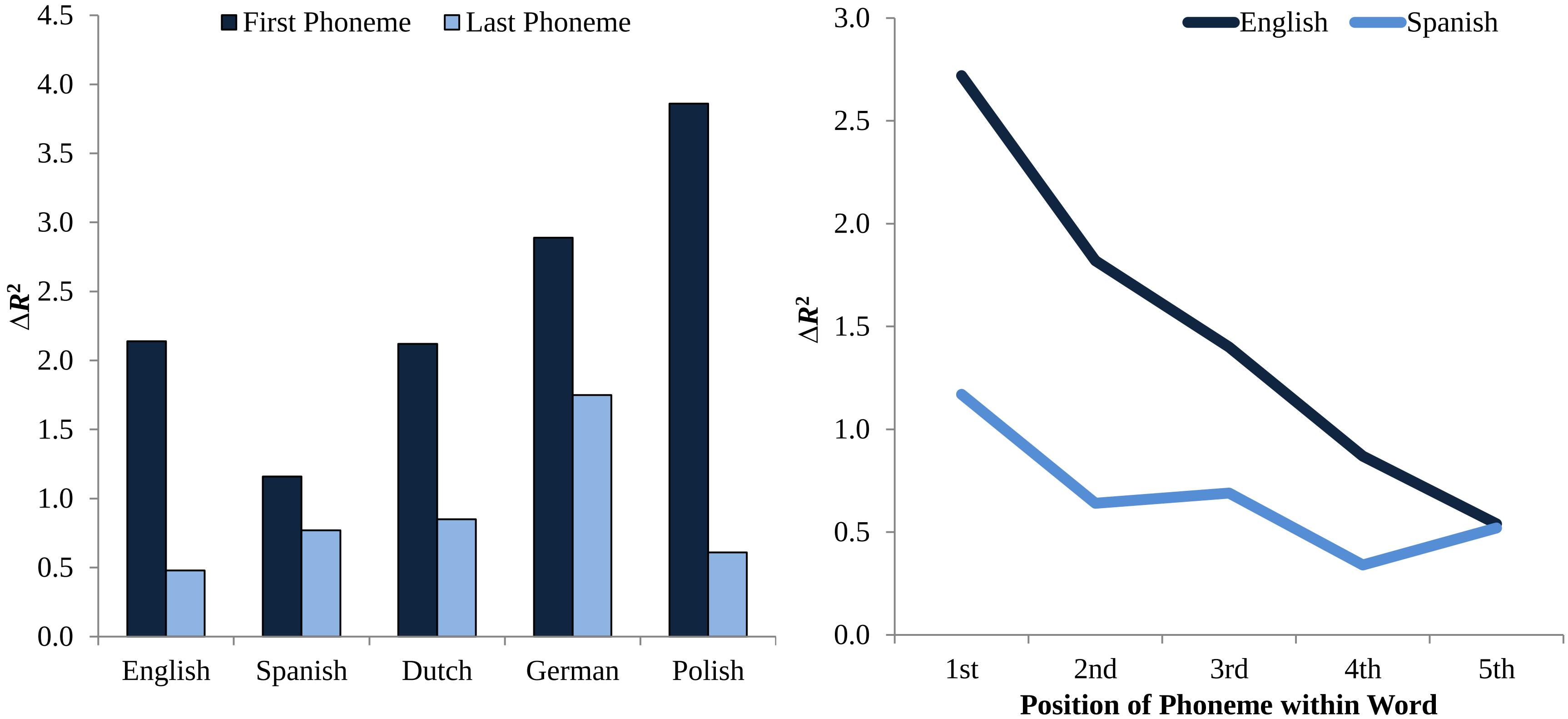
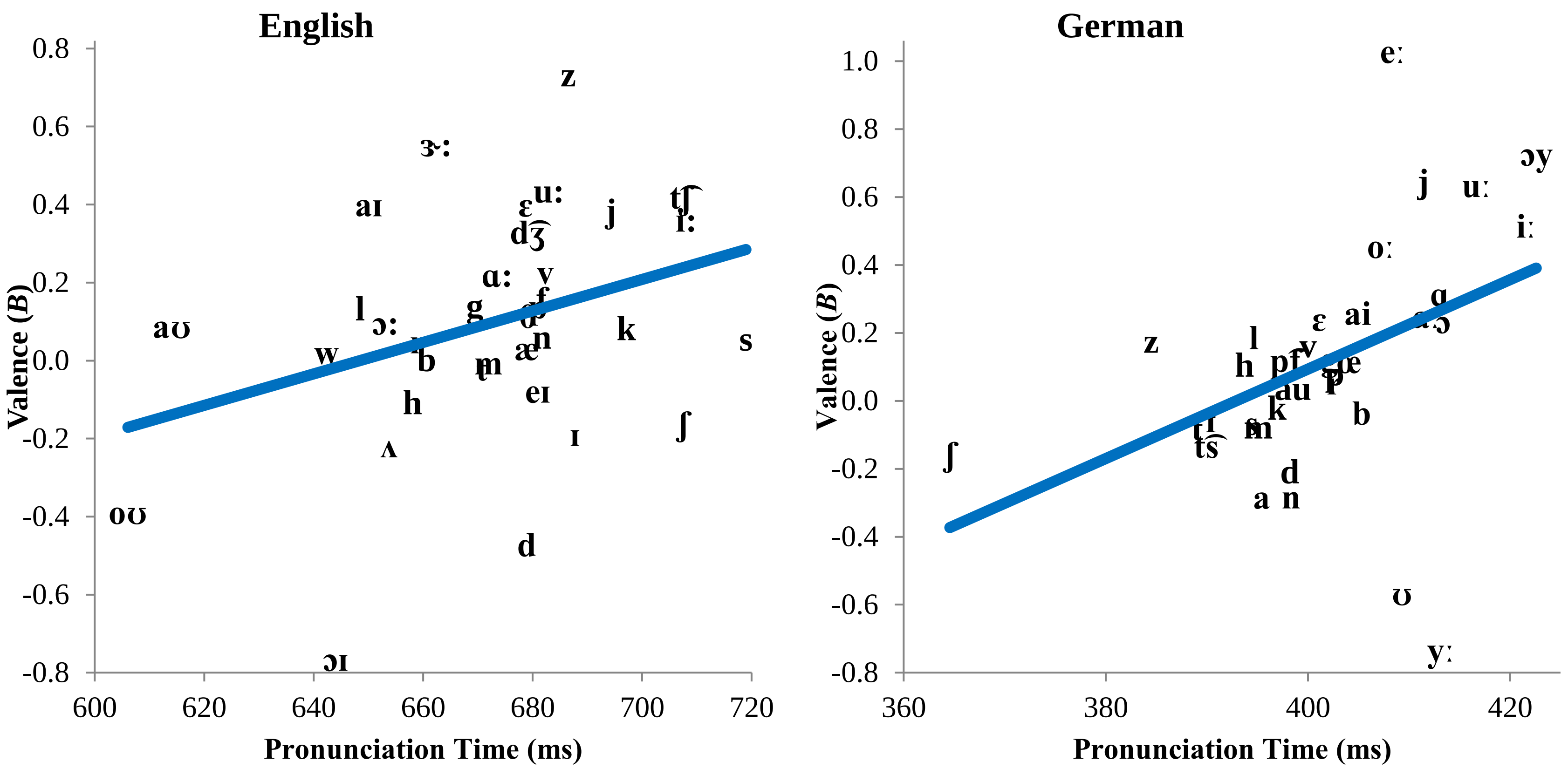


Fig. 4. Negative phonemes are pronounced more rapidly than positive phonemes. Results are illustrated for phonemes in English (r = .34) and German (r = .42) because these are the only languages for which sufficiently large and overlapping datasets of pronunciation latencies and emotion ratings are currently available. Note that one outlying phoneme is excluded from the English illustration; when it is included, the effect is stronger (r = .55).



Supplementary Table 1. Contribution of phonemes to predicting valence in English. Significant predictors ($P < .05$) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	All Words All Phonemes			Monomorph. Words All Phonemes			All Words First Phoneme Only			All Words Last Phoneme Only		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
cent(Arousal)	-0.284	-23.252	**.000**	-0.289	-16.457	**.000**	-0.286	-23.529	**.000**	-0.287	-23.496	**.000**
cent(Length)	0.042	2.852	**.004**	0.037	1.553	.121	0.049	3.946	**.000**	0.060	4.601	**.000**
cent(LgCD)	-0.281	-2.931	**.003**	-0.396	-3.360	**.001**	-0.229	-2.399	**.016**	-0.334	-3.469	**.001**
cent(LgWF)	0.672	7.652	**.000**	0.783	7.370	**.000**	0.622	7.106	**.000**	0.716	8.106	**.000**
cent(Ncons)	NA	NA	NA	NA	NA	NA	-0.037	-2.472	**.013**	-0.040	-2.508	**.012**
cent(Nvows)	NA	NA	NA	NA	NA	NA	0.051	2.568	**.010**	0.009	0.413	**.680**
Control Block R^2:		7.82%	**.000**		10.71%	**.000**		7.82%	**.000**		7.82%	**.000**
ɑ:	0.095	2.569	**.010**	0.165	3.001	**.003**	0.202	1.909	.056	0.863	2.109	**.035**
æ	0.048	1.427	.154	0.089	1.772	.076	0.014	0.190	.850	NA	NA	NA
ʌ	-0.025	-1.063	.288	0.113	2.984	**.003**	-0.235	-4.658	***.000***	0.196	2.133	**.033**
ɔ:	0.000	-0.005	.996	0.058	0.804	.421	0.079	0.591	.554	-0.212	-0.734	.463
aʊ	0.080	1.176	.240	0.126	1.135	.257	0.070	0.392	.695	0.466	1.257	.209
aɪ	0.214	5.466	***.000***	0.197	3.050	**.002**	0.382	2.192	**.028**	0.093	0.639	.523
b	-0.017	-0.519	.604	-0.098	-2.031	**.042**	-0.015	-0.333	.739	-0.051	-0.350	.726
t͡ʃ	0.122	2.067	**.039**	0.094	1.182	.237	0.402	3.621	***.000***	-0.137	-1.174	.241
d	-0.188	-7.330	***.000***	-0.212	-4.951	***.000***	-0.489	-11.283	***.000***	-0.070	-1.705	.088
ð	0.279	1.762	.078	0.129	0.570	.569	NA	NA	NA	0.751	1.497	.134
ɛ	0.112	3.491	***.000***	0.184	3.552	***.000***	0.382	5.070	***.000***	NA	NA	NA
ɝ:	0.006	0.204	.838	0.075	1.501	.133	0.536	2.599	**.009**	-0.077	-2.030	**.042**
eɪ	0.023	0.638	.523	0.125	2.018	**.044**	-0.095	-0.482	.630	0.080	0.667	.505
f	0.080	2.312	**.021**	-0.029	-0.518	.605	0.141	2.864	**.004**	0.087	0.718	.473
g	-0.035	-0.859	.390	-0.151	-2.588	**.010**	0.123	1.946	**.052**	-0.282	-2.203	**.028**
h	-0.149	-2.935	**.003**	-0.242	-3.075	**.002**	-0.125	-2.213	**.027**	NA	NA	NA
ɪ	-0.060	-2.390	**.017**	0.105	2.333	**.020**	-0.207	-4.362	***.000***	NA	NA	NA
i:	0.099	3.785	***.000***	0.071	1.644	.100	0.343	1.465	.143	0.040	1.069	.285
d͡ʒ	0.017	0.341	.733	-0.002	-0.021	.983	0.312	3.268	***.001***	-0.228	-2.368	**.018**
k	0.000	0.013	.989	-0.069	-1.900	.057	0.065	1.837	.066	-0.012	-0.260	.795
l	0.029	1.153	.249	-0.076	-2.025	**.043**	0.115	1.883	.060	0.059	1.726	.084
m	0.031	1.136	.256	-0.077	-1.829	.068	-0.023	-0.493	.622	0.032	0.517	.605
n	-0.012	-0.560	.576	-0.048	-1.329	.184	0.043	0.565	.572	-0.054	-1.622	.105
ŋ	0.061	1.243	.214	0.070	0.772	.440	NA	NA	NA	0.053	0.917	.359
oʊ	0.053	1.329	.184	0.193	3.508	***.000***	-0.406	-3.038	**.002**	0.205	2.295	**.022**
ɔɪ	-0.186	-1.701	.089	-0.005	-0.029	.977	-0.783	-1.288	.198	-0.022	-0.086	.931
p	0.026	0.952	.341	0.044	1.072	.284	0.121	3.125	**.002**	0.063	0.858	.391
r	-0.061	-2.540	**.011**	-0.064	-1.716	.086	0.031	0.708	.479	0.152	1.921	.055
s	-0.087	-4.190	***.000***	-0.155	-4.591	***.000***	0.038	1.204	.229	-0.065	-1.648	.099
ʃ	-0.054	-1.256	.209	-0.078	-0.971	.332	-0.178	-1.833	.067	-0.062	-0.553	.580
t	-0.022	-1.027	.304	-0.105	-3.112	**.002**	-0.039	-0.772	.440	-0.028	-0.927	.354
θ	-0.001	-0.007	.994	-0.028	-0.238	.812	0.096	0.699	.485	0.165	1.045	.296
ʊ	0.250	2.982	**.003**	0.224	1.569	.117	NA	NA	NA	NA	NA	NA
u:	0.108	2.316	**.021**	0.163	2.395	**.017**	0.418	0.595	.552	0.225	1.744	.081
v	0.144	3.570	***.000***	0.071	1.082	.279	0.209	2.523	**.012**	0.123	1.562	.118
w	0.066	1.376	.169	0.021	0.293	.770	0.004	0.053	.958	NA	NA	NA
j	-0.007	-0.115	.909	0.040	0.405	.685	0.367	2.450	**.014**	0.285	0.232	.816
z	0.067	1.430	.153	0.132	1.744	.081	0.715	2.561	**.010**	0.176	2.292	**.022**
ʒ	-0.048	-0.367	.714	0.222	1.068	.286	1.020	0.839	.401	0.112	0.342	.732
Phonemes ΔR^2:		1.44%	**.000**		1.18%	**.000**		2.15%	**.000**		0.48%	**.000**

Supplementary Table 2. Contribution of phonemes to predicting valence in English (replication study). Significant predictors (*P* < .05) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	Adelman et al. norms			Common subset (Adelman et al.)			Common subset (Warriner et al.)		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
cent(Arousal)	-0.559	-17.361	**.000**	-0.589	-16.372	**.000**	-0.274	-9.902	**.000**
cent(Length)	-0.021	-0.533	.594	-0.074	-1.564	**.118**	-0.079	-1.601	.110
cent(LgCD)	-0.101	-0.854	.393	-0.461	-2.599	.009	-0.359	-1.945	.052
cent(LgWF)	0.361	3.587	**.000**	0.759	4.950	**.000**	0.733	4.601	**.000**
Vowels	--	--	--	--	--	--	--	--	--
Consonants	--	--	--	--	--	--	--	--	--
Control Block R^2:		14.72%	**.000**		16.68%	**.000**		11.96%	**.000**
ɑː	-0.254	-1.180	.238	-0.275	-1.089	.276	-0.421	-1.601	.110
æ	-0.292	-1.366	.172	-0.375	-1.497	.134	-0.502	-1.930	.054
ʌ	-0.339	-1.626	.104	-0.409	-1.673	.094	-0.514	-2.024	**.043**
ɔː	-0.348	-1.614	.107	-0.344	-1.355	.176	-0.337	-1.280	.201
aʊ	-0.198	-0.935	.350	-0.206	-0.820	.412	-0.205	-0.789	.430
aɪ	0.008	0.042	.967	0.027	0.112	.911	-0.129	-0.521	.602
b	0.036	0.350	.726	-0.004	-0.037	.971	0.061	0.500	.617
t͡ʃ	0.110	0.855	.393	0.224	1.510	.131	0.213	1.381	.167
d	-0.026	-0.264	.791	-0.070	-0.597	.550	-0.036	-0.294	.768
ð	-0.201	-0.876	.381	0.956	1.906	.057	1.688	3.253	***.001***
ɛ	-0.060	-0.280	.780	-0.082	-0.326	.744	-0.218	-0.835	.404
ɜː	0.003	0.014	.989	0.075	0.323	.747	-0.079	-0.329	.742
eɪ	-0.058	-0.273	.785	-0.145	-0.586	.558	-0.249	-0.970	.332
f	0.096	0.902	.367	0.107	0.860	.390	0.169	1.302	.193
g	0.062	0.559	.576	0.069	0.541	.588	0.118	0.893	.372
h	0.022	0.175	.861	0.033	0.228	.820	0.049	0.325	.745
ɪ	-0.085	-0.398	.690	-0.101	-0.403	.687	-0.272	-1.049	.294
iː	0.043	0.204	.838	-0.021	-0.085	.932	-0.101	-0.392	.695
d͡ʒ	0.213	1.600	.110	0.292	1.899	.058	0.165	1.036	.300
k	-0.005	-0.057	.955	0.025	0.226	.821	0.062	0.531	.596
l	-0.050	-0.498	.619	-0.052	-0.450	.652	0.003	0.025	.980
m	0.054	0.532	.595	0.014	0.122	.903	0.161	1.322	.186
n	0.070	0.730	.466	0.123	1.084	.278	0.179	1.523	.128
ŋ	0.112	0.812	.417	0.092	0.565	.572	0.223	1.318	.187
oʊ	-0.044	-0.206	.837	-0.092	-0.367	.714	-0.188	-0.723	.469
ɔɪ	0.159	0.559	.577	0.114	0.353	.724	-0.063	-0.189	.850
p	0.155	1.590	.112	0.173	1.539	.124	0.200	1.714	.087
r	0.058	0.582	.561	0.066	0.562	.574	0.084	0.692	.489
s	0.019	0.202	.840	0.022	0.198	.843	0.138	1.178	.239
ʃ	0.076	0.579	.563	0.097	0.635	.526	0.137	0.865	.387
t	0.032	0.334	.738	0.044	0.401	.688	0.049	0.423	.672
θ	0.279	2.004	**.045**	0.424	2.529	**.012**	0.403	2.310	**.021**
ʊ	-0.007	-0.026	.980	0.002	0.006	.995	0.059	0.181	.856
uː	-0.034	-0.158	.874	-0.085	-0.341	.733	-0.228	-0.882	.378
v	0.392	2.859	**.004**	0.378	2.398	**.017**	0.309	1.887	.059
w	0.064	0.566	.571	0.102	0.757	.449	0.178	1.272	.203
j	0.275	1.575	.115	0.244	1.137	.256	0.297	1.329	.184
z	0.250	1.852	.064	0.413	2.499	**.013**	0.489	2.847	**.004**
ʒ	-0.165	-0.157	.876	-0.062	-0.057	.955	0.835	0.740	.459
Phonemes ΔR^2:		2.27%	**.001**		2.83%	**.000**		2.57%	**.004**

Supplementary Table 3. Contribution of phonemes to predicting valence in Spanish. Significant predictors ($P < .05$) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	All Phonemes			First Phoneme Only			Last Phoneme Only		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
cent(Arousal)	-0.742	-72.110	**.000**	-0.744	-72.048	**.000**	-0.757	-73.633	**.000**
cent(Length)	-0.176	-5.193	**.000**	-0.166	-6.219	**.000**	-0.148	-5.766	**.000**
cent(LgCD)	0.760	5.917	**.000**	0.836	6.504	**.000**	0.667	5.116	**.000**
cent(LgWF)	-0.151	-1.599	.110	-0.207	-2.203	**.028**	-0.080	-0.833	.405
cent(Ncons)	NA	NA	NA	0.187	7.044	**.000**	0.130	5.101	**.000**
cent(Nvows)	NA	NA	NA	0.226	7.305	**.000**	0.253	8.440	**.000**
Control Block R^2:		29.49%	**.000**		29.49%	**.000**		29.49%	**.000**
a	0.004	0.254	.800	0.076	2.599	**.009**	-0.057	-3.012	**.003**
b	0.135	3.599	***.000***	0.163	4.406	***.000***	NA	NA	NA
β˕	0.081	3.016	**.003**	1.044	1.879	.060	-0.069	-0.108	.914
t͡ʃ	0.155	2.175	**.030**	0.495	3.810	***.000***	NA	NA	NA
d	-0.279	-8.879	***.000***	-0.382	-10.320	***.000***	NA	NA	NA
ð˕	-0.056	-2.409	**.016**	NA	NA	NA	0.133	2.468	**.014**
e	0.018	1.042	.297	0.083	2.462	**.014**	0.074	2.466	**.014**
f	-0.001	-0.043	.965	0.114	2.298	**.022**	0.805	1.445	.148
g	0.177	2.878	**.004**	0.265	3.731	***.000***	NA	NA	NA
ɣ˕	0.001	0.040	.968	0.421	2.289	**.022**	1.057	3.144	***.002***
ʝ	0.066	0.681	.496	-0.108	-0.541	.589	NA	NA	NA
i	0.030	1.667	.096	-0.131	-3.423	***.001***	0.226	1.817	.069
j	-0.158	-5.929	***.000***	-0.679	-0.611	.541	NA	NA	NA
ɲ	0.054	0.629	.529	0.177	0.159	.874	NA	NA	NA
k	0.035	1.594	.111	0.014	0.505	.614	0.735	2.464	**.014**
l	-0.004	-0.179	.858	0.094	1.607	.108	0.045	0.940	.347
ʎ	0.017	0.235	.814	-0.230	-1.029	.303	NA	NA	NA
m	-0.062	-2.699	**.007**	-0.069	-1.773	.076	0.562	1.821	.069
n	0.058	2.982	**.003**	0.023	0.325	.745	0.041	1.238	.216
ŋ	-0.371	-6.485	***.000***	NA	NA	NA	NA	NA	NA
o	0.002	0.128	.898	-0.064	-1.156	.248	-0.119	-6.477	***.000***
p	-0.031	-1.281	.200	0.028	0.908	.364	0.757	1.798	.072
ɾ	0.078	4.615	***.000***	NA	NA	NA	0.183	7.490	***.000***
r	-0.035	-0.978	.328	0.050	1.314	.189	NA	NA	NA
s	-0.101	-5.248	***.000***	-0.072	-1.880	.060	-0.165	-2.815	**.005**
t	-0.027	-1.348	.178	-0.119	-2.721	**.007**	NA	NA	NA
θ	0.053	2.033	**.042**	-0.089	-1.226	.220	0.161	1.468	.142
u	-0.039	-1.492	.136	-0.135	-1.339	.181	0.434	1.350	.177
w	0.011	0.218	.827	-1.140	-1.450	.147	NA	NA	NA
x	0.094	2.391	**.017**	0.290	3.533	***.000***	-0.385	-0.489	.625
z	-0.082	-1.169	.242	0.414	0.527	.598	NA	NA	NA
Phonemes ΔR^2:		1.40%	**.000**		1.15%	**.000**		0.77%	**.000**

Supplementary Table 4. Contribution of phonemes to predicting valence in Dutch. Significant predictors ($P < .05$) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	All Phonemes			First Phoneme Only			Last Phoneme Only		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
cent(Arousal)	-0.028	-1.430	.153	-0.035	-1.784	.075	-0.030	-1.505	.132
cent(Length)	-0.031	-1.219	.223	-0.021	-0.960	.337	-0.015	-0.638	.524
cent(LgCD)	-0.545	-3.891	**.000**	-0.533	-3.827	**.000**	-0.520	-3.671	**.000**
cent(LgWF)	0.658	5.203	**.000**	0.652	5.178	**.000**	0.639	4.994	**.000**
cent(Ncons)	NA	NA	NA	0.019	0.768	.442	-0.001	-0.031	.976
cent(Nvows)	NA	NA	NA	0.047	1.477	.140	0.037	1.137	.255
Control Block R^2:		2.79%	**.000**		2.79%	**.000**		2.79%	**.000**
ɛ:	-0.056	-0.198	.843	NA	NA	NA	NA	NA	NA
œ:	1.080	1.045	.296	1.132	1.097	.273	NA	NA	NA
ə	0.012	0.327	.743	NA	NA	NA	-0.008	-0.213	.832
d͡ʒ	0.061	0.155	.877	0.057	0.146	.884	NA	NA	NA
ø:	-0.110	-0.979	.328	-0.834	-0.808	.419	-0.008	-0.016	.987
ʉ	-0.090	-1.113	.266	NA	NA	NA	NA	NA	NA
ɒ:	0.583	0.976	.329	NA	NA	NA	NA	NA	NA
a:	-0.017	-0.331	.740	0.012	0.076	.939	0.281	1.945	.052
ɑ	-0.033	-0.589	.556	-0.388	-3.171	**.002**	NA	NA	NA
b	-0.004	-0.084	.933	-0.010	-0.176	.860	NA	NA	NA
d	-0.031	-0.627	.531	-0.134	-1.876	.061	NA	NA	NA
e:	0.032	0.531	.595	0.464	2.619	**.009**	0.205	0.764	.445
ɛ	0.055	0.929	.353	0.134	0.819	.413	NA	NA	NA
f	0.076	1.249	.212	0.406	3.744	***.000***	-0.145	-1.454	.146
g	0.326	1.240	.215	0.305	0.590	.555	NA	NA	NA
ɣ	-0.108	-1.333	.183	NA	NA	NA	NA	NA	NA
h	0.010	0.146	.884	-0.03	-0.377	.706	NA	NA	NA
i:	0.064	1.195	.232	0.538	1.473	.141	0.086	0.649	.517
ɪ	-0.060	-0.926	.354	-0.032	-0.185	.853	NA	NA	NA
j	0.228	2.485	**.013**	0.073	0.366	.715	0.034	0.162	.871
k	-0.002	-0.051	.960	0.020	0.386	.700	-0.004	-0.067	.946
ɛi	0.068	0.943	.346	0.263	1.136	.256	0.474	2.089	**.037**
l	0.015	0.388	.698	0.016	0.205	.838	-0.045	-0.838	.402
œy	0.077	0.836	.403	0.324	1.085	.278	-0.26	-0.613	.540
m	-0.005	-0.107	.915	-0.108	-1.505	.132	0.131	1.557	.119
ɑu	0.226	1.879	.060	0.089	0.258	.797	1.239	1.194	.232
n	0.099	2.302	**.021**	-0.052	-0.439	.660	0.087	1.115	.265
ŋ	-0.189	-2.347	**.019**	NA	NA	NA	-0.273	-2.317	**.021**
o:	-0.023	-0.424	.671	0.260	1.353	.176	0.366	2.186	**.029**
ɔ	-0.189	-3.228	***.001***	-0.494	-5.272	***.000***	NA	NA	NA
p	0.118	2.625	**.009**	0.122	1.934	.053	0.053	0.601	.548
r	-0.017	-0.491	.623	0.033	0.442	.659	0.062	1.088	.277
s	-0.105	-2.976	**.003**	-0.089	-2.061	**.039**	-0.083	-1.467	.142
ʃ	0.311	2.320	**.020**	0.345	1.562	.118	0.413	1.589	.112
t	0.045	1.365	.172	0.114	1.637	.102	-0.009	-0.229	.819
u:	-0.008	-0.099	.921	0.269	0.369	.712	-0.565	-1.632	.103
v	0.045	0.794	.427	0.088	1.352	.176	NA	NA	NA
ʋ	-0.194	-3.276	***.001***	-0.092	-1.095	.274	-0.062	-0.383	.702
x	-0.012	-0.302	.763	0.084	1.228	.220	-0.063	-0.924	.355
y:	0.049	0.459	.646	0.360	0.781	.435	0.066	0.110	.912
z	0.050	0.733	.463	0.066	0.743	.458	NA	NA	NA
ʒ	0.612	2.448	**.014**	0.403	0.873	.383	NA	NA	NA
Phonemes ΔR^2:		2.32%	**.000**		2.14%	**.000**		0.82%	.056

Supplementary Table 5. Contribution of phonemes to predicting valence in German. Significant predictors ($P < .05$) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	All Phonemes			First Phoneme Only			Last Phoneme Only		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
Arousal	-0.925	-31.330	**.000**	-0.900	-30.737	**.000**	-0.920	-31.449	**.000**
Length	-0.043	-1.214	.225	-0.005	-0.179	.858	-0.037	-1.374	.169
Frequency	0.139	13.779	**.000**	0.126	12.762	**.000**	0.136	13.626	**.000**
Consonants	NA	NA	NA	0.024	0.753	.451	0.059	1.691	.091
Vowels	NA	NA	NA	0.053	1.407	.159	0.093	2.496	**.013**
Control Block R^2:		27.58%	**.000**		27.58%	**.000**		27.58%	**.000**
a	-0.012	-0.207	.836	-0.281	-3.415	***.001***	NA	NA	NA
ɛː	0.125	0.562	.574	NA	NA	NA	NA	NA	NA
œ	0.365	0.899	.369	0.119	0.114	.909	NA	NA	NA
ə	-0.052	-1.099	.272	NA	NA	NA	-0.058	-1.123	.262
œ̃ː	1.258	1.193	.233	NA	NA	NA	1.187	1.125	.261
d͡ʒ	1.040	1.703	.089	1.433	1.929	.054	NA	NA	NA
øː	0.050	0.222	.824	NA	NA	NA	-0.716	-0.677	.498
ɒː	1.258	1.684	.092	NA	NA	NA	1.774	1.676	.094
p͡f	0.000	0.001	.999	0.123	0.420	.674	-0.531	-1.123	.262
t͡s	-0.125	-1.727	.084	-0.130	-1.096	.273	0.309	2.031	**.042**
6ː	1.956	1.853	.064	NA	NA	NA	NA	NA	NA
aː	0.070	1.117	.264	0.252	0.926	.355	0.362	2.135	**.033**
b	0.007	0.112	.911	-0.033	-0.457	.648	NA	NA	NA
au	0.073	0.794	.427	0.040	0.263	.792	-0.037	-0.142	.887
d	-0.121	-1.702	.089	-0.206	-1.685	.092	NA	NA	NA
eː	0.181	2.718	**.007**	1.032	4.261	***.000***	0.322	1.005	.315
ɛ	0.074	1.108	.268	0.244	2.074	**.038**	NA	NA	NA
f	-0.021	-0.376	.707	-0.055	-0.825	.410	-0.119	-0.732	.464
g	-0.051	-0.811	.418	0.126	1.383	.167	NA	NA	NA
h	0.082	0.918	.359	0.108	1.036	.300	NA	NA	NA
iː	0.227	4.053	***.000***	0.518	1.098	.272	0.702	3.983	***.000***
g	0.276	4.025	***.000***	0.317	1.414	.158	NA	NA	NA
j	0.701	3.270	***.001***	0.649	2.052	**.040**	NA	NA	NA
t͡ʃ	0.070	0.148	.883	0.469	0.447	.655	-0.332	-0.445	.657
k	0.024	0.465	.642	-0.018	-0.239	.811	0.095	0.928	.353
l	-0.090	-2.006	**.045**	0.189	2.214	**.027**	-0.011	-0.160	.873
m	-0.082	-1.462	.144	-0.074	-1.002	.317	0.047	0.387	.699
n	-0.126	-2.893	**.004**	-0.279	-2.288	**.022**	-0.160	-3.742	***.000***
ŋ	0.078	0.881	.379	NA	NA	NA	0.325	3.156	**.002**
oː	0.047	0.748	.455	0.458	1.508	.132	0.308	1.481	.139
ɔ	0.137	1.685	.092	0.237	1.006	.315	NA	NA	NA
p	0.022	0.350	.726	0.083	0.943	.346	-0.204	-0.981	.327
ɑ̃ː	1.180	1.931	.054	NA	NA	NA	NA	NA	NA
r	-0.058	-1.389	.165	0.056	0.584	.559	0.028	0.439	.661
s	-0.145	-2.391	**.017**	-0.062	-0.205	.837	-0.171	-1.790	.074
ʃ	-0.075	-1.030	.303	-0.153	-2.041	**.041**	0.004	0.019	.985
t	-0.043	-1.042	.297	-0.078	-0.915	.360	0.085	1.528	.127
uː	0.354	4.143	***.000***	0.637	1.358	.175	0.503	1.062	.288
ʊ	0.104	1.346	.179	-0.564	-3.656	***.000***	NA	NA	NA
v	0.019	0.259	.796	0.166	1.801	.072	NA	NA	NA
ai	0.239	2.959	**.003**	0.261	1.382	.167	0.083	0.358	.720
x	-0.084	-1.070	.285	NA	NA	NA	0.134	1.086	.278
ɔy	0.135	0.925	.355	0.730	1.201	.230	0.026	0.042	.966
yː	0.105	0.624	.532	-0.730	-2.385	**.017**	NA	NA	NA
ʏ	0.465	2.610	**.009**	NA	NA	NA	NA	NA	NA
z	-0.042	-0.611	.542	0.179	1.853	.064	NA	NA	NA
ʒ	0.570	1.507	.132	2.678	3.605	***.000***	NA	NA	NA
Phonemes ΔR^2:		2.68%	**.000**		2.73%	**.000**		1.82%	**.000**

Supplementary Table 6. Contribution of phonemes to predicting valence in Polish. Significant predictors (*P* < .05) are in bold font, and those that remained significant after Bonferroni correction are also italicized.

	All Phonemes			First Phoneme Only			Last Phoneme Only		
Predictor	***B***	***t***	***p***	***B***	***t***	***p***	***B***	***t***	***p***
cent(Arousal)	-0.289	-6.945	**.000**	-0.275	-6.673	**.000**	-0.301	-7.127	**.000**
cent(Length)	-0.094	-0.994	.320	-0.020	-0.486	.627	-0.019	-0.460	.646
cent(LgCD)	2.555	2.437	**.015**	3.015	2.883	**.004**	3.026	2.829	**.005**
cent(LgWF)	0.204	5.699	**.000**	0.202	5.671	**.000**	0.212	5.800	**.000**
cent(Ncons)	NA	NA	NA	0.006	0.115	.909	0.003	0.053	.958
cent(Nvows)	NA	NA	NA	0.143	2.827	**.005**	0.098	1.935	.053
Control Block R^2:		4.70%	**.000**		4.70%	**.000**		4.70%	**.000**
a	0.075	1.842	.066	0.074	0.555	.579	0.006	0.115	.908
b	-0.241	-3.400	***.001***	-0.133	-1.273	.203	NA	NA	NA
d	-0.170	-2.659	**.008**	0.078	0.648	.517	NA	NA	NA
d͡z	-0.076	-0.355	.722	-0.026	-0.045	.964	NA	NA	NA
d͡ʐ	0.406	1.040	.298	-0.213	-0.323	.747	NA	NA	NA
d͡ʑ	-0.058	-0.406	.684	0.316	1.418	.156	NA	NA	NA
ɛ	0.028	0.544	.586	0.634	2.207	**.027**	-0.041	-0.377	.706
ɛ̃	-0.131	-1.210	.226	NA	NA	NA	0.620	1.603	.109
f	0.092	0.922	.357	0.286	1.501	.133	-0.176	-0.800	.424
g	-0.182	-2.380	**.017**	-0.291	-2.243	**.025**	NA	NA	NA
i	0.045	0.755	.450	0.234	1.173	.241	0.118	0.782	.434
ɪ	0.139	2.796	**.005**	NA	NA	NA	0.091	1.745	.081
j	0.216	3.950	***.000***	0.557	2.536	**.011**	0.112	0.372	.710
k	-0.044	-0.972	.331	-0.050	-0.628	.530	-0.072	-0.766	.444
l	0.072	1.240	.215	0.251	1.846	.065	0.151	0.852	.394
m	-0.055	-0.943	.346	0.145	1.617	.106	-0.183	-0.861	.389
n	-0.056	-1.143	.253	-0.4	-2.843	**.005**	0.061	0.415	.678
ɲ	-0.432	-5.783	***.000***	-0.996	-6.548	***.000***	-0.066	-0.271	.787
o	0.143	3.220	***.001***	-0.163	-1.860	.063	-0.071	-0.613	.540
ɔ̃	0.052	0.447	.655	NA	NA	NA	NA	NA	NA
p	-0.072	-1.404	.160	0.009	0.169	.866	-0.184	-0.723	.469
r	-0.123	-2.567	**.010**	0.108	0.951	.342	0.059	0.434	.664
s	0.033	0.647	.518	0.065	0.790	.430	0.101	0.626	.531
ʂ	0.007	0.061	.951	0.033	0.215	.830	0.117	0.615	.538
ɕ	0.057	0.664	.507	0.239	1.572	.116	0.503	1.057	.291
t	-0.118	-2.596	**.009**	-0.045	-0.407	.684	-0.094	-0.996	.320
t͡s	-0.054	-0.679	.497	0.053	0.251	.802	0.037	0.167	.868
t͡ʂ	0.207	1.673	.094	0.159	0.724	.469	-0.007	-0.021	.984
t͡ɕ	-0.154	-2.873	**.004**	0.076	0.289	.773	-0.058	-1.242	.214
u	-0.074	-1.290	.197	-0.063	-0.491	.623	-0.134	-0.116	.908
v	0.124	2.514	**.012**	0.295	3.857	***.000***	0.198	0.510	.610
w	-0.172	-2.178	**.030**	0.347	1.573	.116	NA	NA	NA
x	0.050	0.420	.675	0.000	0.003	.997	-0.468	-1.781	.075
z	-0.131	-1.955	.051	-0.310	-3.729	***.000***	NA	NA	NA
ʒ	-0.105	-1.156	.248	-0.026	-0.134	.893	NA	NA	NA
ʑ	0.056	0.209	.835	0.371	0.460	.645	NA	NA	NA
Phonemes ΔR^2:		4.28%	**.000**		4.03%	**.000**		0.91%	.703

Supplementary Materials: Testing an Interaction Model

Emotion is experienced on two basic dimensions of valence (negative/positive) and arousal (calming/exciting; Lindquist et al., 2012; Russell, 2003). Our analyses in the main text include arousal as a control factor, thereby isolating the relation between phonemes and valence, because our adaptive view of sound symbolism emphasizes the urgency of negative valence. However, some have argued that valence and arousal have interactive effects, such that highly arousing negative words (e.g., “snake”) are especially prioritized because they represent survival threats (Robinson et al., 2004). Unarousing negative words (e.g., “coffin”), in contrast, are nonthreatening and therefore may not be prioritized. Below we report a basic test of this interaction model, but first we consider the relation between valence and arousal and its implication for model testing.

Valence and arousal are negatively correlated. The relationship is actually U-shaped, with negative and positive stimuli tending to be high in arousal and neutral stimuli tending to be low in arousal. However, the U is not symmetric, so there is a significant linear correlation. Given this relation between valence and arousal, it is invalid to make claims about valence without controlling arousal, and vice versa. This is a critical limitation of prior studies on emotional sound symbolism (e.g., Heise, 1966; Louwerse & Qu, in press; Thorndike, 1945), all of which make claims about valence without controlling arousal. In contrast, by controlling arousal in our analyses, we ruled out this potential confound. Rather than controlling arousal, another approach is to vary it and examine its effects. This is essentially what the interaction model proposes, namely, that the interaction of valence and arousal is more important than valence alone. Studies of word recognition and memory do not support this interaction model, as valence and arousal instead have independent, non-interacting influences (Adelman & Estes, 2013; Estes &

Adelman, 2008-a, 2008-b; Kuperman et al., 2014). Essentially, the theoretical rationale *against* the interaction model is that (a) negative low-arousal stimuli are rare, and (b) there is little cost to over-reacting to those stimuli, so the adaptive behavior is to respond strongly to all negative stimuli regardless of their arousal level.

Nonetheless, to test this interaction model in the context of emotional sound symbolism, we created a new dependent variable representing the interaction of valence and arousal. Specifically, we reverse-coded valence and multiplied it with arousal, so that high scores indicate highly arousing negative stimuli, and then we tested whether phonemes predicted that interaction score. We used first phoneme (English language) because it is the most important from the adaptive perspective, and we used the same control variables as our other analyses. We also added valence as a control factor in order to avoid misattributing the phonemes' influence on valence *per se* to their influence on the interaction of valence and arousal. Although the phonemes significantly predicted the interaction variable as a block, $F(35, 12804) = 2.90$, $P < .001$, the pattern of the eight phonemes that reached significance was not consistent with the conceptual model underlying the interaction hypothesis. If highly-arousing negative items drive effects, then the interaction should be significant in the positive direction for phonemes that are negative in valence on the average. Four of the phonemes that reached significance in the interaction analyses were significant in the valence analysis, and these were associated with negative valence, but of these two had positive interactions and two had negative interactions. Thus, there was minimal evidence for an interaction (i.e., significant at the block level) but not consistently in the predicted direction. As such, this interaction model is not theoretically informative in the manner of the simpler valence-only model described throughout the main text (see also Adelman & Estes, 2013; Estes & Adelman, 2008-a, 2008-b; Kuperman et al., 2014). On the other hand, the significant block-level effect suggests there may be some merit to investigating this alternative model further in future research.

Author Notes

J. S. A. and Z. E. contributed equally to this research. Z. E. gratefully acknowledges the generous support of the Center for Research on Marketing and Services (CERMES) at Bocconi University. We also sincerely thank Marc Brysbaert and Padraic Monaghan for constructive feedback on an earlier draft of this manuscript, and Max Louwerse for sharing the English data from Louwerse and Qu (in press) with us. We also thank Tom Barry, Sabrina Golonka, and Patrick Thompson for assistance with preliminary iterations of this work, which was originally presented at the 31st Annual Meeting of the Cognitive Science Society (2009).