

RESEARCH ARTICLE

Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making

Arash Aryani^{1*}, Markus Conrad², David Schmidtke¹, Arthur Jacobs^{1,3}

1 Department of Experimental and Neurocognitive Psychology, Freie Universität Berlin, Berlin, Germany, **2** Department of Cognitive, Social and Organizational Psychology, University of La Laguna, La Laguna, Spain, **3** Centre for Cognitive Neuroscience Berlin (CCNB), Berlin, Germany

* Arash.Aryani@fu-berlin.de

Abstract

Most language users agree that some words sound harsh (e.g. grotesque) whereas others sound soft and pleasing (e.g. lagoon). While this prominent feature of human language has always been creatively deployed in art and poetry, it is still largely unknown whether the sound of a word in itself makes any contribution to the word's meaning as perceived and interpreted by the listener. In a large-scale lexicon analysis, we focused on the affective substrates of words' meaning (i.e. *affective meaning*) and words' sound (i.e. *affective sound*); both being measured on a two-dimensional space of valence (ranging from pleasant to unpleasant) and arousal (ranging from calm to excited). We tested the hypothesis that the sound of a word possesses affective iconic characteristics that can implicitly influence listeners when evaluating the affective meaning of that word. The results show that a significant portion of the variance in affective meaning ratings of printed words depends on a number of spectral and temporal acoustic features extracted from these words after converting them to their spoken form (study1). In order to test the affective nature of this effect, we independently assessed the affective sound of these words using two different methods: through direct rating (study2a), and through acoustic models that we implemented based on pseudo-word materials (study2b). In line with our hypothesis, the estimated contribution of words' sound to ratings of words' affective meaning was indeed associated with the affective sound of these words; with a stronger effect for arousal than for valence. Further analyses revealed crucial phonetic features potentially causing the effect of sound on meaning: For instance, words with short vowels, voiceless consonants, and hissing sibilants (as in 'piss') feel more arousing and negative. Our findings suggest that the process of meaning making is not solely determined by arbitrary mappings between formal aspects of words and concepts they refer to. Rather, even in silent reading, words' acoustic profiles provide affective perceptual cues that language users may implicitly use to construct words' overall meaning.



OPEN ACCESS

Citation: Aryani A, Conrad M, Schmidtke D, Jacobs A (2018) Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. PLoS ONE 13 (6): e0198430. <https://doi.org/10.1371/journal.pone.0198430>

Editor: Emmanuel Manalo, Kyoto University, JAPAN

Received: April 21, 2017

Accepted: May 18, 2018

Published: June 6, 2018

Copyright: © 2018 Aryani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Human language has generally been considered to be entirely symbolic in that words convey meaning through conventional and arbitrary links to concepts they refer to [1]. From this

perspective, phonemes (i.e. the speech sounds that constitute words) have no inherent semantic content nor have they any stand-alone contribution to words' meaning. Nevertheless, even a naïve reader—without prior knowledge of such literary devices as cacophony or euphony—would experience how, for instance, in Poe's verse “. . .Hear the loud alarum bells—Brazen bells!—What tale of terror, now, their turbulency tells!” [2], the explosive consonant /t/ and other harsh and discordant sounds (e.g. hissing sibilants /s/ and /z/) evoke a feeling of “terror” provoked by “brazen” bells.

Within literary studies, many have noted that poetry achieves much of its affective aesthetic impact through sound manipulation, and that phonological structure has a semantic function beyond the decorative [3–5]. In a similar fashion, swear words usually possess specific phonological patterns that can potentially amplify the negative emotional response that they mean to evoke [6]. Looking at the famous seven words listed by American comedian George Carlin that “you can never say on television” [7] reveals that all of these words contain voiceless stops (/t/ and /k/) or hissing sibilants (/s/ and /ʃ/), which are fortis consonants, articulated with greater oral pressure and relatively higher muscular force compared to their lenis counterparts.

However, despite the fact that influential linguists and experimental psychologists throughout the last century promoted the idea that the sound of a word may have a synchronic, productive effect on overall meaning construction [8–10], the notion of the arbitrariness of the linguistic sign [1] has generally dominated research on human language.

More recently, a growing body of research challenges the idea of absolute arbitrariness by providing evidence for non-arbitrary sound-to-meaning correspondences (see [11–13] for reviews) including some universal patterns across various languages of the world [14]. These results assign a supplementary function to sound-to-meaning correspondences that structure vocabulary [15,16] and play an important role for both phylogenetic language evolution [16–18] and ontogenetic language development [18,19]. Nonetheless, despite the increasing number of studies examining sound-to-meaning associations, to the best of our knowledge, there has been no empirical study examining whether specific properties in the sound of a real word play a part in contributing to its overall meaning. With the present study, we aimed at addressing this research question. By focusing on the ‘affective meaning’ of words, and by providing reliable quantitative measures for ‘affective sound’ of words, we investigated how the sound of a word potentially contributes to its meaning as perceived and evaluated by the listener. A further goal of this study was to explore the affective acoustic cues and their underlying phonetic features that may implicitly influence language users when evaluating words' affective meaning.

Motivation for the present study

Our approach was motivated by a number of limitations evident in previous work. Experimental research based on behavioral data has hitherto merely investigated the links between some selective, rather isolated attributes of meaning (e.g. the physical size of the referent) and some aspects of sound (e.g. intrinsic pitch of vowels) mainly by using nonword stimuli (see [20] for supporting a graded relationship between sound and meaning, and [21], for an evolutionary perspective on the phenomenon). Such approaches exhibit three major limitations that we aimed to address in the present study.

The first limitation relates to the focus on semantic effects of phonemes in nonwords instead of natural words. Such studies are motivated by the fact that natural words in a language are linked to predetermined semantic concepts that are automatically activated during word recognition. In order to disentangle the effect of phonology from that of semantics, the majority of previous studies therefore relied on nonword stimuli usually used in a forced-

choice paradigm thus limiting the generalizability of the results to real words. For instance, the phonemes /ɑ/ and /ɪ/ when used in experimentally manipulated nonwords—as in “mal” and “mil” in the seminal study by Sapir [9]—have repeatedly been suggested to denote big and small objects, respectively [11,12]. However, in a natural language like English, they appear in the corresponding semantic concepts in the opposite way: /smɑl/ and /bɪg/. This begs the question to what extent the results of these studies can be linked to natural language processing and whether the assumed quality of phonemes has, if any, effects on the evaluation of meaning for real words.

A second issue relates to the problem of deciphering the likely cause(s) of sound-to-meaning correspondences. Proposals on non-arbitrariness of language distinguish between two types of motivations for such sound-meaning mappings [12]: Iconicity, which is based on perceptual similarities between sound and meaning (e.g. onomatopoeia), versus systematicity which is based on statistical regularities in language that link specific patterns of sound to specific semantic or grammatical concepts [22,23]. Besides some familiar and straightforward examples of iconicity—such as onomatopoeic words—research in this field still faces the question of whether existing findings on the relationship between sound and meaning are caused by specific distributions of phonemes in a language (i.e. systematicity), or by perceptual qualities that phonemes inherently convey (i.e. iconicity). The phonaestheme /sn-/ appearing as an initial sound cluster in many English words related to ‘mouth’ or ‘nose’ may serve to illustrate this dilemma [24]. In this case, there has been no empirical support showing whether there is a specific (nasal) quality in the sound of /sn-/ that is linked with the concepts of ‘mouth’ or ‘nose’, or rather the organization of the vocabulary is designed in a way that this specific sound cluster over-proportionally appears in words that are related to these concepts.

The third and presumably most important issue is that the operationalization of meaning in this field of research has so far been restricted to only some selective aspects of sensorimotor information (e.g. shape, movement). The role of affect as a most basic human experience shaping the learning, representation, and processing of language [25–29] has been surprisingly neglected. Indeed, affective dimensions of words, in particular, valence and arousal, are essential features defining a two-dimensional semantic space allowing for a very basic and potentially the most relevant distinction between different concepts; as empirically established by semantic differential [30]. In an attempt to provide a quantitative measure for words’ meaning, Osgood [30] defined 100 different lexical dimensions and asked participants to allocate the meaning of words for each dimension in an experiential continuum definable by a pair of polar terms (e.g. soft/hard, long/short, angular/rounded). Factor analyses conducted on the wide variety of verbal judgments indicated that most of the variance was accounted for by three major semantic dimensions: The two primary dimensions of ‘valence’ and ‘arousal’, and a third, less strongly-related dimension (in terms of the explained variance) of ‘dominance’ or ‘control’ [31]. Therefore, these factors have been considered basic dimensions of the semantic space within which the meaning of any concept can be specified.

Moreover, the expression and perception of affective states are fundamental aspects of human communication [32,33] that have been proposed as the original impetus for language evolution; with mimetic vocalization of emotional sounds supposedly allowing early hominids to efficiently share biologically significant information [32,34,35]. Therefore, we would expect the effect of iconicity to be most evident in the communication of affect and in the relationship between words’ affective sound (i.e. how emotionally words sound) and words’ affective meaning (i.e. their position in the bi-dimensional affective space of lexical valence and arousal). Thus iconicity can serve as an interface for accomplishing the need to map linguistic form to human affective experience as a vital part of meaning making.

An embodied view on affective meaning

It is important to consider that the notion of “affective meaning” may not be shared by all theories on linguistic meaning. Our approach in this work is based on an embodied view of language which proposes that meaning is grounded in behavior (perception and action) and neural circuitry of the producer or the interpreter of linguistic signs [25,28,36–40]. Ultimately, part of the meaning of any utterance is its effect on the (physical and emotional) well-being of the person saying or hearing it, and everything that matters is represented in each individual person’s brain and its neurophysiological systems. Presumably, the most fundamental such system is affect: in order to make meaning, we need to know what object/event in our environment requires us to react with alert or to keep calm, to approach or to withdraw. Moreover, the ability to distinguish between such affective contexts or reactions is linked to attention systems that select specific sensory input for further processing, and also to motor systems that select specific actions for output. Both systems (i.e. sensory and motor) provide crucial information for the construction of meaning by language users. Findings on the role of affective meaning in modulating various cognitive processes, such as learning, memory, attention or language processing, [25,26,28,41] support the idea that affective meaning is intertwined with other lexico-semantic aspects and has an essential and basic contribution to the process of meaning making.

The present study

We addressed the above-mentioned problems apparent in previous research by focusing on the affective meaning of real words and investigated whether participants were implicitly influenced by words’ sound while giving a rating on emotion expressed in words’ meaning.

Specifically, we aimed to challenge the established notion that assigning (affective) meaning to words is merely determined by words’ semantic content and by an associative and per se arbitrary relationship between the signifier (sound image) and the signified (concept)—as encouraged by a leading principle of modern linguistics [1]. Instead, we propose that the overall affective meaning of a word is, in addition to the word’s semantic content, co-determined by inherent qualities of the signifier and by the percept derived from words’ acoustic-phonetic features (i.e. the *affective sound*). Note that our use of the term affective sound in this paper refers exclusively to phonological constituents of words themselves and not to speaker-related issues such as intonation or how a word is spoken.

Our main hypothesis is motivated by research on nonverbal emotional vocalization and, in particular, emotional prosody which has shown that the emotional significance of a sound can be detected, and hence be integrated with higher-order cognition, even when the attentional focus is not directly on the emotional cues of the sound [42–45]. Such emotional cues have been shown to be engaged even in silent reading by means of cross-sensory input from the visual cortex into the auditory cortex and affective regions in the brain [46,47]; as put forward by theories of embodied cognition and perceptual simulation. On the other hand, phonemes and their combinations (as in words) are characterized by a number of acoustic features that overlap with those that modulate emotional vocalization and emotional prosody (e.g. sound formants, sound intensity). Therefore, the specific sound profile of any word in a language can theoretically be attributed to a specific emotion as perceived by the listener. We thus hypothesized that the process of affective lexical evaluation—as higher order cognition—is influenced by words’ phonology: that is, their phonologically recoded neuronal representation of the acoustic features corresponding to phonological word forms [48–51].

We used a large-scale normative database including rating values for affective meaning of words that were evaluated by at least 20 subjects/item. In line with Osgood’s semantic

differential [30], such databases usually contain two types of ratings. The first component concerns emotional valence going from unpleasant to pleasant. For instance, words such as “murder”, “poison”, and “virus” are commonly rated as “very unpleasant” whereas “freedom”, “love”, and “life” lie on the other extreme end. The second type of ratings addresses the degree of emotional arousal ranging from excited (e.g. “nightmare”, “sex”, and “courage”), to calm (e.g. “health”, “massage”, and “peace”). Using these ratings as a measure of words’ affective meaning, we tested the null hypothesis (H0) that explicit evaluations of affective meaning solely reflect written words’ semantic content, against the alternative hypothesis (H1) that phonological word forms also contribute to valence and arousal ratings so that a statistically significant portion of their variance can be accounted for by words’ acoustic features. For instance, harsh-sounding words might make people feel more aroused so that they implicitly give a higher arousal rating, even though they are instructed to only focus on the lexico-semantic aspect of words.

In order to test H1, we chose a computational approach that employs signal averaging to amplify the potential effect of sound on meaning. Subsequently, we quantified a *Phonological Affective Potential* (PAP) of words, separately for arousal (PAP_{aro}) and for valence (PAP_{val}), both estimates representing the influence of words’ affective sound on their affective meaning. The goal of this work is to examine the psychological reality of the PAP, and to test whether the PAP is linked to (and derived from) words’ acoustic features that we extracted from their spoken forms after synthesizing them (study 1). Note that according to H0 the amount of variance in the PAP that depends on words’ acoustic features should not be greater than chance level and therefore statistically not significant. Next, we tested the association of the PAP with words’ affective sound assessed in two behavioral studies using different methods. We first asked participants to rate the affective sound of printed words while suppressing words’ meaning (study 2a), and second, we employed auditory presented pseudowords and acoustic models to predict words’ affective sound based on their acoustic features (study 2b). We then compared the PAP with these two independent measures of affective sound to test for their potential associations. Finally, in order to identify the perceptual cues potentially underlying the effect of implicit sound on meaning, we separately tested the relationship between words’ acoustic features and words’ PAP and the two independent measures for words’ affective sound.

Although previous studies provided first affirmative support for the affective potential of some phonological units [52–56], ours is the first study demonstrating the psychological reality of phonemes’ affective potential and the contribution of words’ implicit sound to meaning making for real words and across a language’s lexicon.

Quantifying the *Phonological Affective Potential* (PAP)

Material

The Berlin Affective Word List [57] (BAWL) was used as a normative database containing a representative mass of 2694 German words that has been cross-validated in various empirical studies regarding experiential, behavioral, and neurobiological levels of analysis [40]. The BAWL includes words from different classes (nouns, verbs, and adjectives) that were selected based on the following main criteria: to include a) the most frequently used German words, b) as many words as possible with an apparent relation to affect regardless of whether this would result in more or less extreme values of valence and arousal, and c) a critical mass of theoretically neutral words. As a consequence, the BAWL contains a relatively elevated percentage of emotion-laden words. However, valence and arousal values of words in the BAWL are spread across the entire range of both valence and arousal in order to make it an optimal tool for

selecting verbal material for all kinds of research questions on affective language processing. Henceforth, we refer to these valence and arousal values representing words' affective meaning as *Affective Meaning Ratings*.

Method

We operationalized H1 by assigning two statistical components to each of the *Affective Meaning Ratings* in the BAWL (separately for arousal and valence): for a word composed of phonemes $W_i [ph_1, ph_2, \dots, ph_n]$, the rating value was considered to reflect a first component stemming from an explicit evaluation of the word's *Affective Semantic Content*, and a second phonological component, the PAP (*Phonological Affective Potential*), which reflects the contribution of the affective potential of phonemes to the total rating value. Assuming a simple additive model, the rating for a given word in the database can thus be modelled as the sum of these two components plus an error term ϵ

$$\begin{aligned} \text{AffectiveMeaning Rating}(W_i) \\ = \text{Affective Semantic Content}(W_i) + \text{PAP}_i[ph_1, ph_2, \dots, ph_n] + \epsilon \end{aligned} \tag{1}$$

We estimated the PAP by averaging the potential affective effects of all phonemes regardless of their position in the word. The PAP thus can be hypothesized as a function of the *Phonological Affective Value* (PAV) of each phoneme: that is, the contribution of the affective quality of each single phoneme to the total rating value for each word:

$$\text{PAP}_i[ph_1, ph_2, \dots, ph_n] = \text{Mean} [\text{PAV}(ph_1), \text{PAV}(ph_2), \dots, \text{PAV}(ph_n)] \tag{2}$$

In order to quantify the $\text{PAV}(ph_i)$, we considered it to be a signal masked by 'noise': that is, the *Affective Semantic Content* (W_i) and the error term in Eq 1. We thus attempted to minimize the effect of noise while increasing the signal-to-noise-ratio (SNR) by averaging. That is, for each phoneme, we calculated the average rating values of words that contain this phoneme (Fig 1A). This way, the *Affective Semantic Content* (W_i) and the error term ϵ nearly cancel out, and the average ratings are approximately associated with the potential contribution of each single phoneme (i.e. the PAV) to the rating value.

Corpus preparation

In order to have an adequate number of repetitions and hence improve the SNR, we chose only those phonemes with a frequency of appearance higher than 30 in the database (mean frequency = 322). This led to the exclusion of 120 words that contained phonemes with a lower frequency, including those that are not a part of German phonology. Overall 12 phonemes were excluded, seven with a frequency of seven to 30 (ʒ, ε:, |:, j, pf, tj, /) and five which did not belong to German phonology: /:, dʒ, æ:, ð:, ã:.

Calculation of PAPs

For a given word W_k composed of n phonemes $\{Ph_1, Ph_2, \dots, Ph_n\}$ with a rating R_k (valence or arousal), we first defined a membership function α as follows:

$$\left\{ \begin{aligned} Ph_i \in W_k &\Rightarrow \alpha(Ph_i, W_k) = 1 \\ Ph_i \notin W_k &\Rightarrow \alpha(Ph_i, W_k) = 0 \end{aligned} \right\}$$

Potential (for valence or arousal) for a given word W_k will then be:

$$PAP(W_k) = \frac{\sum_{i=1}^n \alpha(Ph_i, W_k) \times \frac{\sum_{j=1}^N \alpha(Ph_i, W_j) \times R_j}{\sum_{j=1}^N \alpha(Ph_i, W_j)}}{\sum_{i=1}^n \alpha(Ph_i, W_k)}$$

Note that all of the following studies were conducted for the remaining number of 2574 words for which the PAP was calculated.

Study 1: Relating words’ acoustic features and PAP

If the PAP of written words is somehow linked to emotional cues present in their phonological forms, we would expect it to be related to the acoustic features of that form.

Method

To test the above relationship, we synthesized the words and extracted their acoustic features, focusing on a total of 11 features that are known to modulate emotional vocalization: fundamental frequency (f0; mean), sound intensity (mean and standard deviation), spectral center of gravity (mean), standard deviation of the spectrum, and sound formants (F1, F2, F3; means and bandwidths) [45,58–60].

It is worth pointing out that we deliberately opted for synthesizing the words rather than using a human speaker in order to prevent any undesired emotional prosody that might result from words’ affective meaning: Human speakers tend to pronounce words with a prosodic intonation—independently from phonological content—consistent with words’ meaning. By synthesizing the words, we distinctly separated our dependent and independent variables: PAPs (PAP_{aro} and PAP_{val}) on the one side, and acoustic features on the other side. Although the artificial nature of a synthesized voice could diminish the goodness of acoustic models, a positive result would all the more support the effects in question.

Synthesizing and acoustic analyses

We synthesized the words using the eSpeak [61] as front-end to the male voice de4 from MBROLA [62] which consists of a speech synthesizer; based on the concatenation of diphones, and of diphone databases. We abstained from the use of larger synthesis units (such as whole words or phrases as used in Variable Unit Concatenation systems) to avoid the potential effect of words’ affective content on speakers’ prosody as discussed above [63]. Words were synthesized in a fixed carrier sentence, *Das Wort . . . wird oft verwendet* (“The word . . . is often used”). The rate of speech was set at 120wpm (words per minute). All spoken words were checked for intelligibility by two male native speakers (not otherwise involved in the study). Importantly, the speakers were not provided with the word list so that they had no expectations about the words’ identity [64]. Both speakers agreed on the intelligibility of all words: speaker1 marked four words and speaker2 marked seven words as poorly synthesized; however, they found all words still intelligible. We extracted the acoustic features of words using the speech analysis software PRAAT [65]. We extracted the mean of fundamental frequency f0 (time step = .01, min = 75Hz, max = 300Hz), the mean and standard deviation of intensity (time step = .01), and the mean and the bandwidth of the first three formants F1, F2 and F3

(time step = .01) from the spectral representation of the sound. Finally, the spectral centroid (spectral center of gravity) and the standard deviation of the spectrum were computed on the basis of fast Fourier transformations (time step = .01, min pitch = 75Hz, max pitch = 300Hz).

Results and discussion

We computed two multiple linear regression models to predict the PAP_{aro} and the PAP_{val} using the above distinctive acoustic features as regressors ($N = 11$). The acoustic analyses reported next were carried out on all of the 2574 words in the database. The results are summarized in Fig 1B and S2 Table. For both arousal and valence, PAPs were significantly predicted by the distinctive acoustic variables (both $P_s < 0.0001$), the variance accounted for being 27.9% for the PAP_{aro} and 23.7% for the PAP_{val} (both R^2 adjusted).

Words in the database that are derived from the same stem or root morpheme (e.g. 'terror' and 'terrorize') are likely to have both similar phonological structure and similar semantic content. This could potentially bias the relationship between PAP and the way words sound. Hence, to ensure unbiased estimation, we selected all monosyllabic words from the database ($N = 289$) and repeated the above analysis steps (including new calculations of PAPs) for this subset comprising only monomorphemic words for which stem repetition was precluded. This time, the 11 acoustic variables accounted for 29.3% of the variance in PAP_{aro} and 26.6% in PAP_{val} (both R^2 adjusted, $P_s < 0.0001$). The successful outcome and even larger portion of explained variance corroborates our previous results and validates the method used to uncover the effects of phonological units. We expected to obtain a better approximation for PAV (and consequently PAP) when the number of phonemes in a word is reduced, as is the case for monosyllabic words.

By showing that a considerable part of the variance in PAPs depends on the acoustic features of the spoken word forms, we could reject H_0 stating that PAPs are a mere product of chance. Instead, H_1 was supported: acoustic features of phonemes significantly co-determined words' affective ratings even when they are visually presented (and silently read). This suggests that the contribution of phonological units to the ratings of words' affective meaning—as reflected in the PAPs—emerges from a representation of acoustic properties of words in spoken form. We take this as a first support for the validity or psychological reality of the effect in question.

Having shown that PAPs of written words are significantly associated with the acoustic profile of their spoken form, we next asked whether this association is based on the words' affective sound as assessed in two independent ways (study 2a and 2b).

The following studies were approved by the ethics committee of the Freie Universität Berlin and were conducted in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants gave their consent (in written form for study 2a, and online for study 2b) prior to participating in the study.

Study 2a: Measuring words' affective sound via rating

Stimuli

The stimuli were the 2574 words from the BAWL used in the previous analyses.

Participants

A total of 272 participants were recruited by flyers, email contacts, and Facebook posts, who then rated the words either for valence or for arousal. Of these, 135 participants (82 females, age = 21.3 ± 4.6) rated exclusively for arousal and 137 participants (92 females, age = 23.6 ± 2.9)

for valence. Participants were mostly students from the Freie Universität Berlin who received either psychology course credit or 5 Euros for their participation. All participants reported normal or corrected-to-normal vision and were native German speakers.

Procedure

Words were presented visually. A very similar set of instructions to those used to rate the words' affective meanings [57] was applied here, with one minor modification. Participants were instructed that they would have to suppress the meaning of words and only pay attention to their sound, and this was repeatedly emphasized through the instruction process. We also incorporated the self-assessment manikins (SAM) that were used in the ANEW study [66]. Words were randomly divided into 8 different lists each of which included about 335 items. Words were then rated on both affective sound of valence and affective sound of arousal by different groups of participants in order to exclude the possibility of mutual influence between valence and arousal ratings. The affective sound of valence was rated on a 5-point scale ranging from -2 (*sehr negativ* / “very negative”) through 0 (*neutral* / “neutral”) to +2 (*sehr positiv* / “very positive”). The 5-point affective sound of arousal scale ranged from 1 (*sehr beruhigend* / “very calming”) to 5 (*sehr aufregend* / “very exciting”). The items were randomly presented to minimize primacy or recency effects. On average, the tasks were completed in approximately 25 minutes.

Analysis

Each word was rated by an average of 19.7 participants for valence and 20.4 participants for arousal. In order to assess the degree of agreement among raters, the Interclass Correlation (ICC) was computed for both arousal and valence ratings. Results showed a higher value for arousal (ICC = 0.43) than for valence (ICC = 0.31), indicating a rather poor degree of agreement.

Even though participants were asked to only focus on the affective *sound* of words, their ratings were likely “contaminated” by words' semantic content, since semantic representations are automatically activated during word recognition. This was evident in the correlations between our ratings of affective sound and the original *Affective Meaning Ratings*: $r = 0.32$ for arousal and $r = 0.22$ for valence. To eliminate the undesired effect of words' *Affective Semantic Content* from our ratings of affective sound, we opted for the most conservative approach. We first regressed the PAPs on the *Affective Meaning Ratings*—separately for arousal and valence—and used the residuals as a statistical estimate for words' *Affective Semantic Content* (cf. Eq 1). In a next step, we regressed the estimate for words' *Affective Semantic Content* on our ratings of affective sound and used the z-transformed residuals of this regression as independent measures of words' affective sound. This way, the potential effect of *Affective Semantic Content* was partialled out of rating values of affective sound. The substantially weaker correlations between these “decontaminated” residuals and the original *Affective Meaning Ratings* validated our method: $r = 0.1$ for arousal, and $r = 0.04$ for valence. These two decontaminated residuals (for arousal and valence) were then used as estimates of the words' affective sound. In the following, we refer to these two measures as *Affective Sound Ratings*; in short: AS-R_{aro} (for arousal) and AS-R_{val} (for valence).

Results and discussion

The correlations between AS-R_{aro} and AS-R_{val} on the one hand, and PAP_{aro} and PAP_{val} on the other hand, were highly significant: $r = 0.5$, for arousal, and $r = 0.25$, for valence (both $P_s < 0.0001$). A similar analysis was performed for the subset of monosyllabic words ($N = 289$): AS-R_{aro} and

AS- R_{val} were also significantly correlated with the corresponding PAP_{aro} : $r = 0.46$, and PAP_{val} : $r = 0.32$, respectively (both P s < 0.0001).

These results indicate that the contribution of phonological units to words' affective meaning ratings (i.e. PAPs) is associated with the affective sound of these words, thus providing further support for the psychological reality and the affective nature of PAPs.

Study 2b: Predicting words' affective sound via acoustic models

In the previous study, the poor ICC values suggested that subjective judgments about the affective sound of a word while trying to suppress its meaning can be a difficult task. In this study, we therefore aimed to provide a new measure of affective sound by using meaningless pseudowords that would allow participants to better focus on the sound. We therefore generated and presented pseudowords in auditory form and collected ratings of their affective sound. By extracting the acoustic features of pseudowords and using them as predictors we developed acoustic models capable of predicting the variation in the ratings. Such independent models can then be applied to any word-like item in auditory form to predict its affective sound solely based on its acoustic features, including the real words from the previous studies. Note that since the pseudowords had to be presented to and rated by human subjects, for this task—unlike in Study 1—we used a human voice rather than a synthesizer to generate naturally sounding stimuli and to prevent potential distortion effects of sound peculiarity. However, as pseudowords lack semantic content there was no concern about the influence of meaning on emotional prosody as in the study 1.

Stimuli

To generate pseudowords representative for the phonotactics of German, we used the Wuggy algorithm [67] which generates pseudowords that match a given word template in sub-syllabic structure and transition frequencies, thus obeying a language's phonotactic constraints. Since the pseudowords had to be spoken and rated, to avoid obscurity we restricted the list of word templates to those having up to three syllables and 10 letters. We then chose the first 1500 most frequent nouns from CELEX [68]. For each word, we adapted the program to generate five pseudoword alternatives using Wuggy's default setting. Candidate pseudowords which differed in fewer than two letters (whether added, deleted or substituted) from the nearest real word were excluded due to their similarity to real words (Coltheart distance = 1). For words with more than one remaining pseudoword alternative, the one with a highest Levenshtein distance [69] was selected. The list of pseudowords was checked for pseudohomophones and a too high similarity to real words. Thus, 187 items were excluded: for example, *beim* (similar to the short form of "*bei dem*" = for something), *absads* (similar to the word "*Absatz*" = paragraph). In addition, because of phonotactic problems mostly caused by illegal or very rare grapheme combinations 190 items were excluded: for example *Weckbevezuz*, *Ymiön*, by two native speakers. The remaining 1123 pseudowords were selected for recording.

A professional male actor was recruited in Berlin, Germany, who was a native speaker of German. He had graduated from professional acting school and was regularly employed in radio, television, and stage work. He was paid to participate. Pseudowords were spoken in a list-like manner to prevent affective prosody and were recorded in the "Leibniz-Zentrum Allgemeine Sprachwissenschaft" in Berlin in a professional sound recording booth using a "Sennheiser MKH20" microphone and "Ultra Gain MIC-2000" preamplifier. The audio signal was recorded using the DAT-recorder "TASCAM DA20MKII" with a sampling frequency of 48 kHz and 16 bits per sample.

Participants

A total of 169 participants were recruited by flyers, email contacts, and Facebook posts, who rated the pseudowords either for valence or for arousal. Of these, 85 participants (52 females, age = 26.7 ± 4.3) rated exclusively for arousal and 84 participants (42 females, age = 27.1 ± 3.8) for valence. Participants had the chance to win one of 10 Amazon coupons which were assigned randomly at the end of the study. All participants were native German speakers.

Procedure

In order to afford a convenient method of sampling that was more representative of the general population, the study was conducted online using the SoSci panel [70]. Adapting the instructions used for the original BAWL ratings for written words, participants were invited to carefully listen to the presented item and evaluate how positive or negative (in the case of valence) and how exciting or calming (in the case of arousal) the pseudowords sounded. During the rating process, a “replay” button was offered to provide participants with the opportunity of repeated listening to each presented item. We also incorporated the self-assessment manikins (SAM) that were used in the ANEW study [66]. Importantly, participants were instructed to give their ratings solely based on the sound aspect of items and not based on their similarity to real words. In order to prevent participants from giving a rating for a similar sounding word, a button labeled “concrete word” was provided next to the rating scale, and participants were instructed to use it in case an item might remind them of a German word. 28 items labelled as “concrete word” by more than 50% of participants were then excluded, leaving 1095 items for further analysis. Pseudowords were randomly divided into 4 different lists, each including about 280 pseudowords. The order of presentation was pseudorandomized for each participant. On average, the task was completed in approximately 15 minutes.

Analysis

Affective ratings were obtained for 1095 pseudowords with 17 ratings per item on average (17.2 for arousal, and 17.5 for valence). We extracted the 11 acoustic features from the spoken pseudowords (see [study 1](#)) and performed two multiple regression models using them as predictors of the ratings separately for arousal and valence. These features accounted for 56.3% of the variance in arousal ratings and 11.2% for valence (both R^2 adjusted, $P_s < 0.0001$, [Fig 2](#)).

Since our ultimate goal was to predict the affective sound of real words, in order to assess how the results of the above models generalize to an independent data set (i.e. real words), we used two-fold cross-validation. The dataset was randomly shuffled into two subsets with equal size one for training and one as a test set, and vice versa. Model accuracy for each run was 57.3% and 52.6% (both R^2 adjusted, $P_s < 0.0001$) for the arousal model, and 10.1% and 9.9% for the valence model (both R^2 adjusted, $P_s < 0.0001$). These are very robust results in terms of explained variance compared to the original models.

Results and discussion

The degree of agreement among raters, compared to the results of study 2a, was considerably higher for both valence (ICC = 0.61) and arousal (ICC = 0.86). The substantial amount of variance accounted for in our regression model for arousal indicates that the affective sound of word-like stimuli could be mapped out in terms of their acoustic cues; a strong evidence for acoustic features to possess affective value on their own. A closer look at the variation in ratings revealed a smaller relative standard deviation for valence (13%) than for arousal (18%), suggesting a lower consensus among participants when rating valence. The considerably

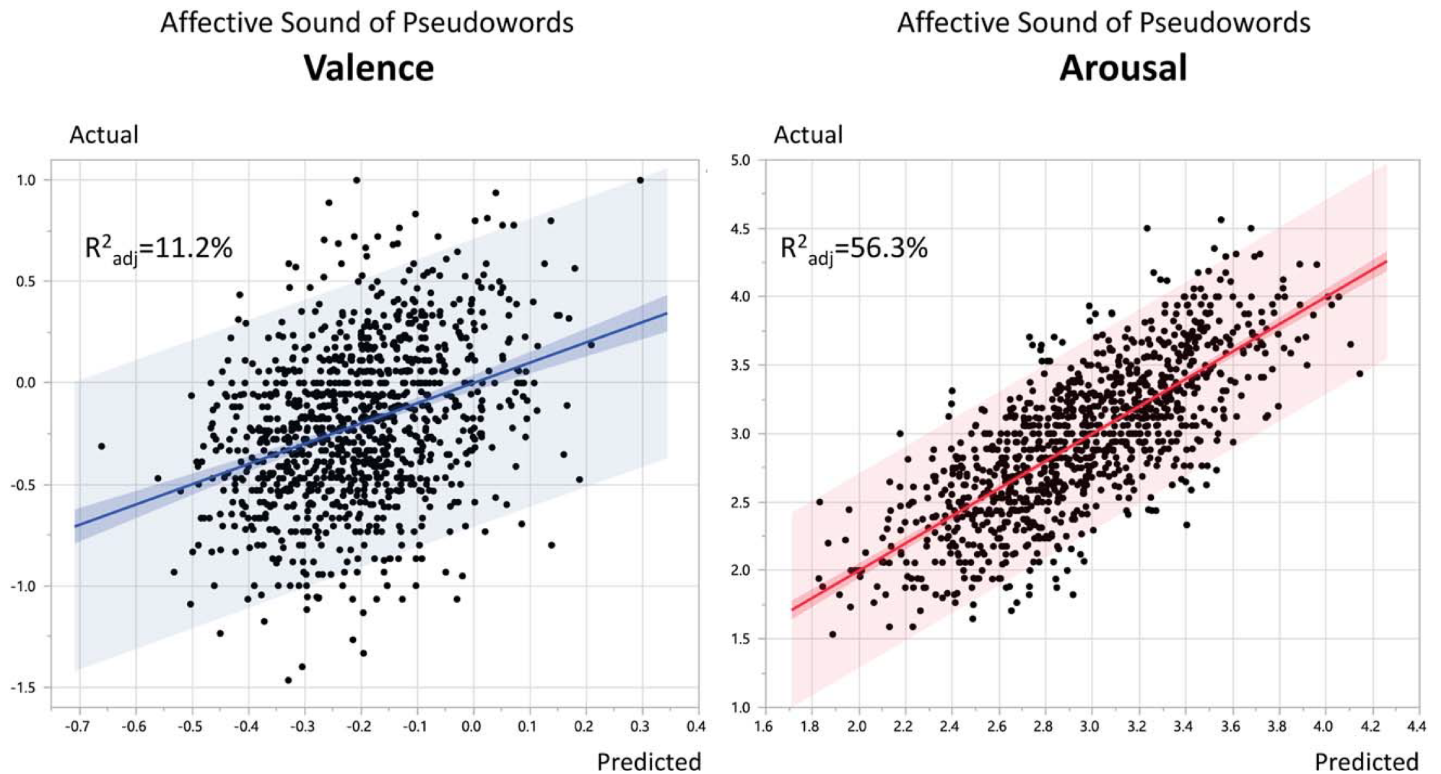


Fig 2. Acoustic features of pseudowords ($N = 11$) significantly predicted the ratings of their affective sound: 11.2% for valence (left) and 56.3% for arousal (right).

<https://doi.org/10.1371/journal.pone.0198430.g002>

higher degree of explained variance for arousal as compared to valence supports the idea that speech sounds primarily signals the sender’s arousal state (and their valence state only to a smaller degree) [59,71]; we will discuss this finding more fully later in this article.

We next took the two acoustic models (i.e. the linear equations in S3 and S4 Tables) resulting from the pseudoword data and applied them to the extracted acoustic features of the words in the database to predict words’ affective sound. We refer to these predicted values for words’ affective sound as *Affective Sound Predicted*; in short: AS- P_{aro} (for arousal) and AS- P_{val} (for valence).

The obtained predicted values for words’ affective sound (i.e. AS-Ps) were then compared with PAP: The AS- P_{aro} and the AS- P_{val} of words were significantly correlated with the PAP $_{\text{aro}}$: $r = 0.47$, and with the PAP $_{\text{val}}$: $r = 0.36$, respectively (both P s < 0.0001). Again, similar results were obtained for monosyllabic words ($N = 289$): AS- P_{aro} was significantly correlated with PAP $_{\text{aro}}$: $r = 0.45$, $P < 0.0001$, and AS- P_{val} was significantly correlated with PAP $_{\text{val}}$: $r = 0.42$, $P < 0.0001$.

These significant associations between words’ PAP and their affective sound of words— independently predicted from acoustic features—add additional support for our H1. In addition to the direct correlation between PAPs and words’ affective sound, as captured by AS-Ps (AS- P_{aro} and AS- P_{val}), we tested the relationship between AS-Ps and those proportions of variance in PAPs that we could account for by means of acoustic features in the first analysis: that is, the predicted values for the PAP $_{\text{aro}}$ and the PAP $_{\text{val}}$ in the first multiple regression models (Study1, Fig 1B) that were calculated with the same acoustic variables as regressors. Results showed high correlations between AS- P_{aro} and the predicted values for PAP $_{\text{aro}}$: $r = 0.88$, and between AS- P_{val} and the predicted values for PAP $_{\text{val}}$: $r = 0.71$ (both P s < 0.0001). This suggests

that the PAPs are based on the same distinctive acoustic features that participants used to evaluate the affective sound of pseudowords, thus, again, providing strong evidence for the association between PAPs and affective sound, and that a significant portion of variance in the ratings of words' affective meaning is due to how words affectively sound.

Furthermore, we tested the reliability of our two different measures of words' affective sound as described in Study 2a and Study 2b, to investigate their consistency in capturing the same concept. For this, we compared the values resulting from these completely independent methods for measuring words' affective sound. Results showed significant correlations between the measure of affective sound based on the direct rating value (i.e. AS-R, Study 2a) and the predicted values of affective sound based on acoustic features (i.e. AS-P, Study 2b) for both arousal: $r = 0.56$, $P < 0.0001$, and valence: $r = 0.49$, $P < .0001$. These results, together with the fact that PAPs are associated with words' affective sound, provide firm support for our H1 stating that the affective meaning of words is shaped by both words' semantic content and (implicit) affective sound.

Analysis of words' acoustic profiles

Having shown a robust association between PAPs and two independent measures of affective sound (AS-R and AS-P), we continued with a more fine-grained analysis and asked whether the underlying acoustic features shaping PAP and both AS-R and AS-P, do so in identical or differential ways for these different measures. Thus, we examined the direct relationships between each acoustic feature and the PAP on the one hand, and our two measures of affective sound (i.e. the AS-R from study 2a and the AS-P from study 2b) on the other.

Method

We constructed acoustic profiles based on the strength and direction of correlations between each of 11 acoustic variables with PAP, AS-R, and AS-P (see Fig 3, see S5 Table for correlation coefficients).

Results and discussion

For the arousal dimension, all single correlations ($N = 3 \times 11$) were highly significant ($P_s < 0.0001$). Notably, the correlations between each acoustic variable and PAP_{aro} were always in the same direction as correlations between this specific acoustic variable and both measures of affective sound for arousal (i.e. the $AS-R_{aro}$, and the $AS-P_{aro}$) resulting in highly similar acoustic profiles for all three measures (Binomial test: $X \sim B(11, 0.5)$, $p(X \geq 11) = 0.0005$). A similar pattern was observed for valence. The PAP_{val} was significantly correlated with seven acoustic variables, and, importantly, these correlations were, again, always in the same direction as for the acoustic variables and both measures of affective sound for valence (i.e. the $AS-R_{val}$, and the $AS-P_{val}$) again resulting in highly similar acoustic profiles for all of three measures (Binomial test: $X \sim B(7, 0.5)$, $p(X \geq 7) = 0.007$).

All correlations in the acoustic profile of arousal remained significant after Bonferroni correction for multiple comparisons. For the acoustic profile of valence, however, the correlation between the third formant (F3) and PAP_{val} did no longer reach statistical significance after Bonferroni correction. But, still, acoustic profiles for all of three measures ($AS-R_{val}$, $AS-P_{val}$, PAP_{val}) remained highly similar (Binomial test: $X \sim B(6, 0.5)$, $p(X \geq 6) = 0.015$).

These results go beyond the simple relationships between the PAPs and the affective sound of words (as captured by AS-P and AS-R); moreover, they show that the acoustic features that underlie PAPs contribute in very similar ways to the perception of words' affective sound. We

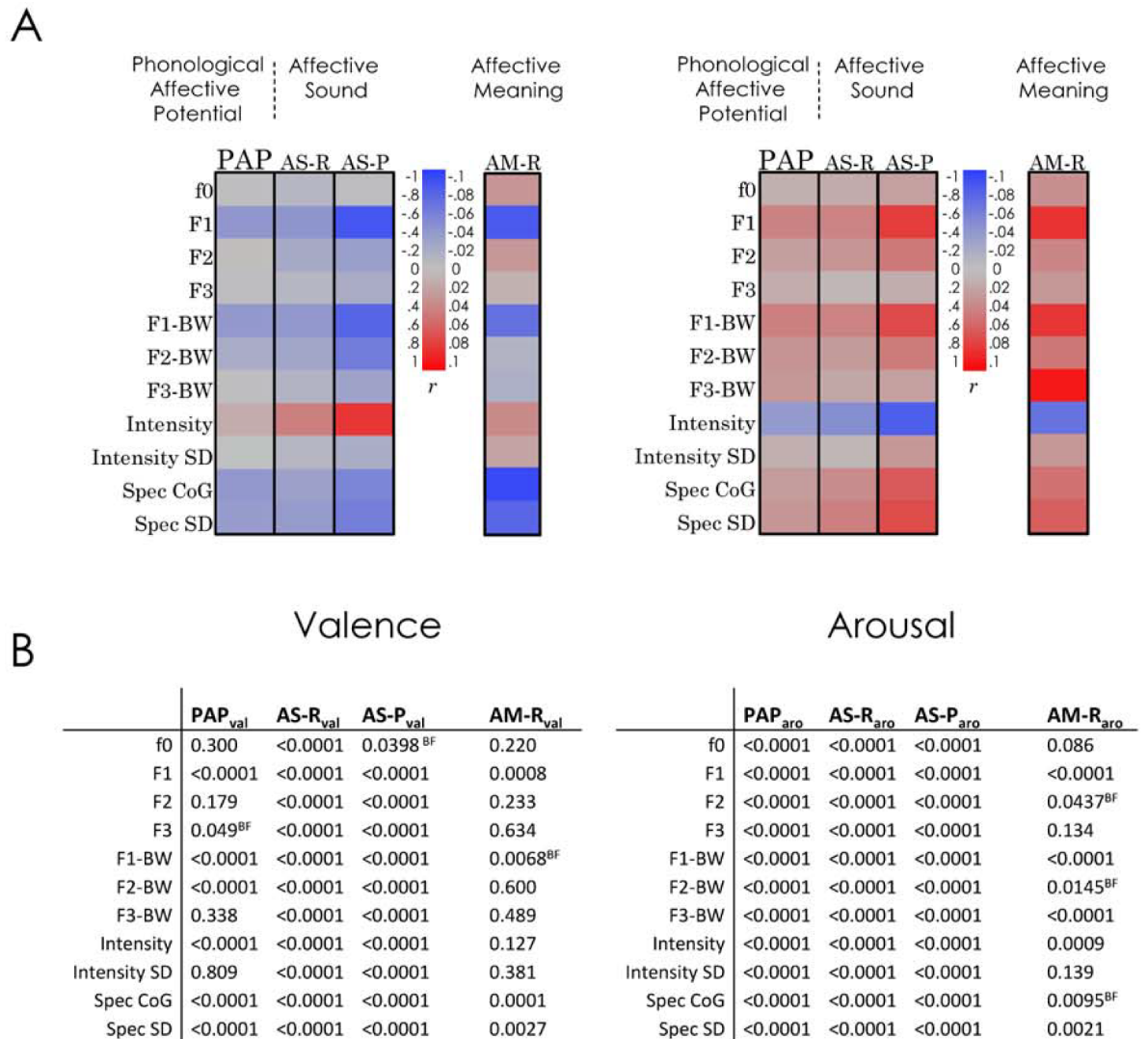


Fig 3. A) Acoustic profiles were constructed (using correlation cell plot) based on the strength and direction of correlations between the estimated effect of words' phonology on the evaluation of their affective meaning (i.e. Phonological Affective Potential: PAP), the two measures of words' affective sound (i.e. Affective Sound-Ratings: AS-R [study 2a], Affective Sound-Predicted: AS-P [study 2b]), and ratings of words' affective meaning (i.e. Affective Meaning-Ratings: AM-R) on the one hand, and 11 acoustic variables on the other hand (left for valence, right for arousal). Acoustic features that significantly correlated with the PAP, AS-R, AS-P, and AM-R always show associations in the same direction, suggesting that acoustic features underlying the affective sound of words contribute in similar ways to the constitution of affective meaning of these words. **B)** The correlation probabilities are shown in the table. Correlations not surviving Bonferroni correction for multiple comparisons are marked with "BF" (Bonferroni Failed). Abbreviations: BW = Bandwidth, SD = standard deviation, Spec = Spectral, CoG = Centre of Gravity, r = correlation coefficient.

<https://doi.org/10.1371/journal.pone.0198430.g003>

interpret this as strong support for PAPs being determined by affective perceptual cues within phonological word forms.

The direct effect of sound on words' affective meaning

Here, we asked whether the contribution of words' (implicit) sound to words' affective meaning can be directly observed at the level of original valence and arousal ratings in the database: that is, before estimating the effect through our statistical operationalization for the PAP. In other words, if an *Affective Meaning Rating* consists of *Affective Semantic Content* and

Phonological Affective Potential (PAP), as formulated in Eq 1, we would expect that the same acoustic features shaping the PAP should be reflected, though to a lesser degree, in *Affective Meaning Ratings*. That is the effect of words' acoustic features on words' affective meaning should be observable directly at the level of *Affective Meaning Ratings*.

Method

We constructed acoustic profiles for *Affective Meaning Ratings* by calculating correlations between *Affective Meaning Arousal-Ratings* (in short AM-R_{aro}) and *Affective Meaning Valence-Ratings* (in short AM-R_{val}) and each of the acoustic variables across all words in the database (N = 2574).

Results and discussion

From the total of 11 acoustic variables, eight variables in the acoustic profile for arousal (five variables after Bonferroni correction) and four in the acoustic profile for valence (three variables after Bonferroni correction) were significantly correlated with AM-R_{aro} and AM-R_{val}, respectively (Fig 3 and S5 Table). Most importantly, those acoustic features that significantly correlated with AM-R_{aro} and AM-R_{val} always showed an association in the same direction as the one between the acoustic features and respective PAP, as well as both measures of words' affective sound (i.e. the AS-R and AS-P). Again, these results support the direct relationship between words' acoustic features and ratings of affective meaning.

Together with our previous findings, these data suggest that the process of meaning making is not solely determined by arbitrary mappings between words' phonology and concepts they refer to. Rather, words' acoustic profiles provide affective perceptual cues that language users implicitly use to construct words' affective meaning.

Acoustic phonetic cues underlying the effect of sound on meaning

Revealing the perceptual acoustic cues likely underlying the effect of implicit sound on affective meaning, we performed further analyses to explore phonetic features potentially causing this effect.

The consistently negative correlations between sound intensity and each of the four arousal-based measures: PAP_{aro}, AS-R_{aro}, AS-P_{aro}, and AM-R_{aro} (Fig 3) deserves a more detailed discussion as arousal usually increases with sound intensity when the latter is experimentally manipulated. Note that all words and pseudowords were spoken with the same loudness, thus differences in sound intensity have to be tracked back to specific phonetic features of the words in the database.

Long vs. short vowels

A closer look at the spectrograms reveals that words with the highest sound intensity tend to include long vowels (e.g. Lohn /l o: n/ "wage", See /z e:/ "lake", see Fig 4A). To systematically examine this potential relationship, we defined a Vowel Length Index as the average vowel length (short = 1, long = 2) over the word's syllables. This Vowel Length Index was significantly correlated with sound intensity across all words in the database: $r = 0.28$, $P < 0.0001$, suggesting a systematic relationship between the two measures. With regard to affective perception, note that long vowels are produced through a release of air from the mouth for an extended period of time which is a behavior similar to slow (vs. rapid) breathing that, in turn, is associated with decreasing (vs. increasing) arousal [72,73]. This relationship between affective states and sound duration is also stressed in the motivation-structural rules hypothesis

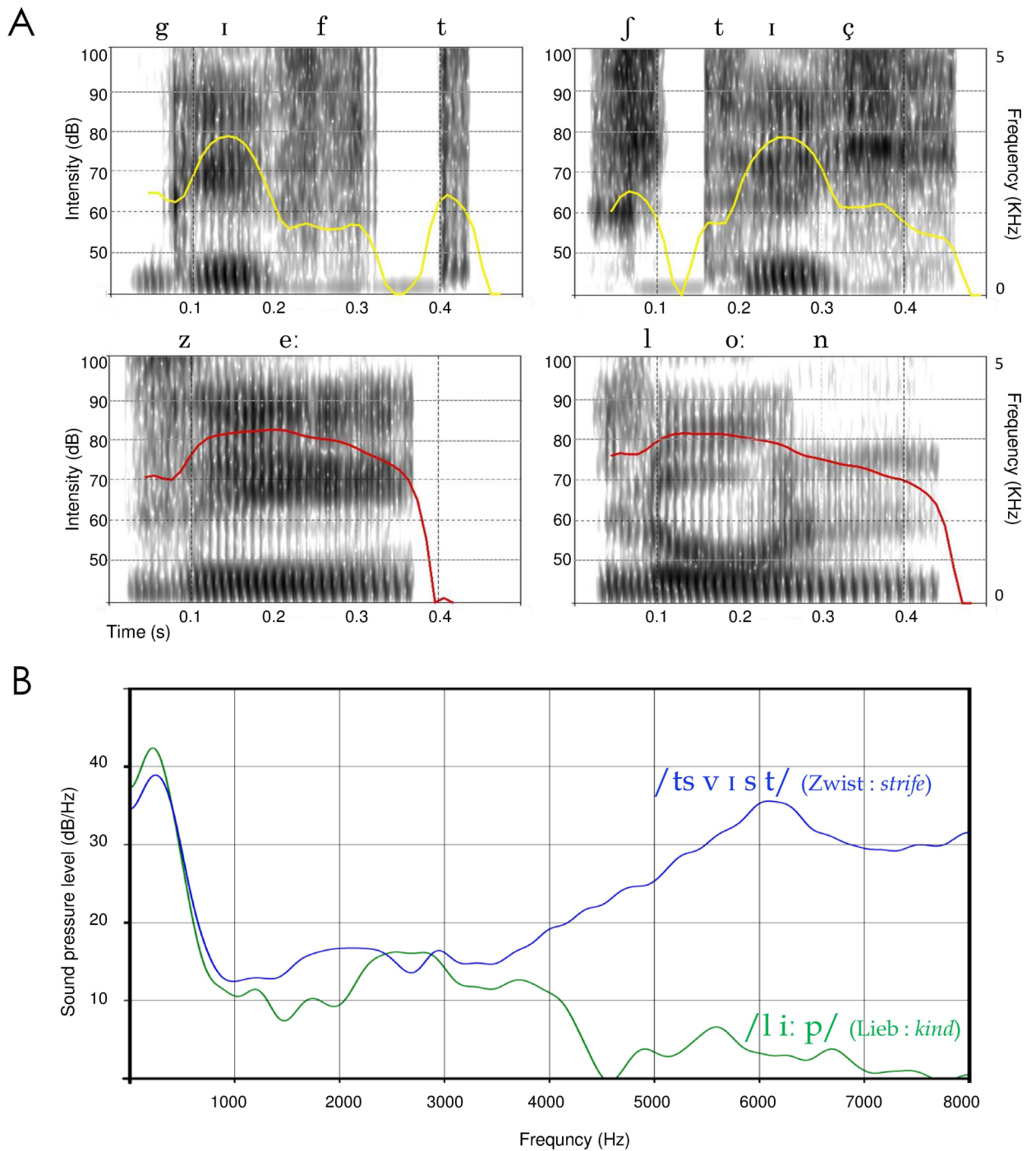


Fig 4. A) The time course of sound intensity for the words “Gift /g I f t/ (*gift*)” and “Stich /ʃ t I ç/ (*stab*)” (top, yellow lines) compared to their counterparts “See /z e:/ (*lake*)” and “Lohn /l o: n/ (*wage*)” (bottom, red lines). Short vowels, plosives, and voiceless consonants (as in “Gift” and “Stich”) possess smaller integrals of sound energy, whereas sustained high amplitude (see red lines) results in larger sound intensity. This relationship between phonetic features and sound intensity, together with the relationship between sound intensity and ‘affective sound’ of words, explains the harsh sound of words containing short vowels, plosives, and voiceless consonants. **B)** Spectral analysis shows that hissing sibilants in a word increase the sound’s center of gravity (i.e. the magnitude-weighted mean of the frequencies present in the signal), which makes words including this category of phonemes sound harsh and negative (blue line *Zwist* /ts v I s t/ (*strife*) vs. green line *Lieb* /l i: p/ (*kind*)).

<https://doi.org/10.1371/journal.pone.0198430.g004>

[74] stating that calls produced by mammals in aggressive circumstances, termed *barks* or *grunts*, are generally of shorter duration than those produced in appeasement contexts. On the other hand, at the spectrogram level, the sustained high amplitude for long vowels causes a larger integral of energy for the whole sound envelope leading to the negative correlation between arousal and sound intensity (see Fig 4A). Note also that the variation of intensity of sound over time (Intensity-SD) accordingly displays a positive correlation with arousal.

A comparison between the PAV of short vowels and their long counterparts (see S1 Table) revealed the same pattern: each of the short vowels was perceived as “more arousing” than its long counterpart: $PAV_{\text{aro}}(/a/) > PAV_{\text{aro}}(/a:/)$, $PAV_{\text{aro}}(/ɔ/) > PAV_{\text{aro}}(/o:/)$, $PAV_{\text{aro}}(/ʊ/) > PAV_{\text{aro}}(/u:/)$, $PAV_{\text{aro}}(/ɪ/) > PAV_{\text{aro}}(/i:/)$, $PAV_{\text{aro}}(/ɛ/) > PAV_{\text{aro}}(/e:/)$. A very similar pattern was revealed for valence values calculated for short and long vowels; with short vowels being more “negative” than their long counterparts—except for the short and long vowels /ɔ/ and /o:/, for which all calculated values were very close to zero.

In addition, short vowels tend to be followed by more consonants (i.e. more complex consonant clusters) than long vowels, and this complexity of subsequent consonant clusters may also hold partly responsible for the observed correlation between vowel length and arousal.

Voicing

Another phonetic feature directly related to sound intensity is ‘voicing’. Voiced consonants are accompanied by vocal cord vibration that leads to an increase in sound energy compared to their voiceless counterparts. In order to explore the relationship between voicing and the affective sound of words, we defined a phonetic cue based on the relative proportion of voiced consonants to all consonants in a word. This phonetic cue of voicing was significantly correlated with sound intensity, $r = 0.38$ ($P < 0.0001$), and also with both measures of affective sound: $AS-R_{\text{aro}}$: $r = -0.51$, and $AS-R_{\text{val}}$: $r = 0.49$, as well as $AS-P_{\text{aro}}$: $r = -0.57$, and $AS-P_{\text{val}}$: $r = 0.62$, (all P s < 0.0001).

These results indicate that voiceless consonants sound on average more arousing and negative than voiced consonants, which, in turn, appear to make words sound softer and more pleasing.

Plosive consonants

Among words with the lowest sound intensity, many include plosive consonants (e.g. Gift /gɪft/ “gift”, Stich /ʃtɪç/ “stab”). The interruption and explosive release of the air stream in the pronunciation of plosive sounds can be associated with a higher level of arousal, but at the same time, during a stop closure, there is very little acoustic energy. This may explain the lower level of sound intensity (and a higher level of arousal at the same time) for words that include this type of phonemes (see Fig 4A).

Similar to voicing, we defined a phonetic cue indicating the relative proportion of plosive consonants to all consonants in a word. This phonetic cue was significantly correlated with sound intensity: $r = -0.26$ ($P < 0.0001$) and with both affective sound measures, $AS-R_{\text{aro}}$: $r = 0.2$, and $AS-R_{\text{val}}$: $r = -0.16$, as well as $AS-P_{\text{aro}}$: $r = 0.17$, and $AS-P_{\text{val}}$: $r = -0.19$, (all P s < 0.0001), reflecting that while plosives reduce sound energy, they also play a significant role in making the sound (moderately) more negative and arousing.

Hissing sibilants

In addition to sound intensity and in line with previous findings on vocal expression of emotion [58–60], first formant (F1) and spectral centroid (CoG) appeared to be the dominant features explaining the largest part of variance in words’ affective sound, showing a significant

effect even at the level of direct ratings for words' affective meaning. A larger high-frequency energy and raising of the first formant are typical characteristics of hissing sibilants (alveolar fricatives and affricates, e.g. /s/, /z/, /ʃ/) which are strongly stressed consonants produced by a high-velocity jet of air against the teeth (see Fig 4B). This results in a literally high-arousing hissing sound, which may account for the cross- and paralinguistic use of these sounds for attracting the attention of others (e.g. "psst!") as well as for their prominent deployment in literature as a stylistic device for cacophony.

Similarly, it is the presence of such a hissing sound following a short vowel that makes the small, but striking difference at the phoneme level between two words referring to one and the same concept from a very basic domain of physical human experience, out of which one is considered rather vulgar and rude, while the other seems more childish and polite: 'piss' vs. 'pee'.

General discussion

The present data demonstrate that words' affective meaning, as reflected in evaluative ratings, is co-determined by words' acoustic-phonetic features. Overall, the results of our computational approach and acoustic analyses, together with the data from the behavioral studies, provide strong support for the hypothesis that phonemes possess affective potential based on their spectro-temporal acoustic features known to modulate emotional vocalization. These results emphasize the iconic nature of the relationship between the (implicit) sound of a phoneme and its affective quality on the one hand and affective meaning of words comprising these phonemes on the other.

As outlined in the introduction, with this study we addressed three major issues generally involved in previous research on iconicity. First, by focusing on a representative number of real words—instead of pseudowords—, our novel results improve the understanding of the effect of implicit sound on the process of meaning making for natural words, in particular concerning their affective meaning. We showed that not only specific sound profiles of words have an affective quality, but also that this quality implicitly influences language users in their final emotional judgment about the meaning of words. Secondly, our behavioral studies and acoustic analyses helped to overcome a major limitation of previous work showing that the relation between affective sound and meaning reflects more than just some statistical regularities within the language (i.e. systematicity) to which language users might be sensitive. Rather, our data suggest that the sound shape of words possesses an inherent affective quality (i.e. iconicity) based on acoustic features that are known to modulate nonverbal emotional communication. Finally, investigating the role of affect and affective meaning of words, we moved beyond the narrow focus on single, limited semantic concepts (see also [19]), which enabled us to test sound-meaning correspondences across a wide range of words, representative of the entire lexicon.

Importantly, the iconic affective potential of phonemes (i.e. PAP), as suggested by our results, contributes to the process of affective meaning making even when words are visually presented and silently read. Note that visual word recognition generally involves the activation of phonological codes [48–50] and language users appear implicitly influenced by affective sound of visually presented words when evaluating the affective meaning of these words.

Valence vs. arousal

Overall, our results were generally stronger for arousal than for valence. This finding aligns with a number of studies on the acoustic properties of emotional speech and hence provides support for an "acoustic arousal dimension". That is acoustic speech properties provide vocal

cues to the level of arousal, above that of valence [56,59,60,75]. Reviews of earlier attempts to decode emotional significance from vocal cues commonly failed to identify a set of vocal features that reliably differentiate between the levels of valence. Arousal is generally understood as related to a physiological state of being reactive to a stimulus, and it appears plausible that this could be reflected in the vocal behavior of the sender and thus extend to acoustic features of the speech signal. Valence, on the other hand, involves higher order, cognitive, and evaluative processes that are less likely to be detectable at such a basal sublexical level [56,76–78], thus corresponding less well than arousal to any consistent acoustic mapping.

Phonetic features

Our analyses of acoustic cues revealed some specific phonetic features as potential candidates of carrying the effect of sound on meaning. Short vowels, compared to their long counterparts, can make words sound more negative and arousing. Also, voiceless consonants, hissing sibilants, and—to some extent—plosives, can significantly contribute to making a word more negative and arousing—as our data suggest, both at the level of sound and perceived meaning.

It is worth noting that these phonetic cues may not be universal across different languages, as not all languages display systematic variations of some of the phonetic features that we focused on (e.g. long/short vowels). Users might rely on different phonological/acoustic affective cues in different languages depending on their phonemic inventory and phonotactic rules—to be investigated in future research on the topic.

Measuring the affective sound of words

Our two studies present, for the first time, two different methods for assessing words' affective sound that can be used in future studies investigating the interaction of words' affective sound and meaning. The poor ICC values for the first method (Study 2a) indicate the difficulty of subjective judgments of the implicit sound of a visually presented word independently of its meaning. Even though we attempted to decontaminate these rating values from the effects of semantic content, this method possesses serious limitations and the poor ICC values call for cautious interpretation of these results.

In contrast to the first method, by using pseudoword material in auditory form in the second study (study 2b), we could largely overcome the limitations of the first approach and provide a better way for assessing words' affective sound, as indicated by the considerably larger ICC values for the pseudoword ratings. Thus our approach based on pseudowords may represent a reliable proxy for words' affective sound in future research.

Alternative interpretation

The present approach aimed at describing the relation between words' phonology and affective ratings in most basic ways, but our findings might also fit well into proposals concerning iconicity and the organization of the vocabulary: Rather than reflecting a direct, forward influence of acoustic features on affective ratings, PAVs, determining PAPs for words in our data might instead, reflect the systematic occurrence of specific phonemes in words of specific affective meaning (in terms of arousal and valence levels) across the vocabulary of a language. This is because PAVs are computed as the average of affective ratings of words comprising a given phoneme. In that case, our data establishing close relations between PAP (or PAV) and acoustic features would help explain an apparent systematic distribution of phonemes across the vocabulary as a function of semantic affective values of words: An iconic relation would link affective attributes of the percept or the basic linguistic sign at the phoneme level with affective semantic meaning at the lexical level—adding an internal to the external relation between the

signifier and the signified that would have contributed to the evolution of the vocabulary according to affective iconicity.

Limitations and future research

Our study is the first to demonstrate an association between affective sound and meaning for real words and across a language lexicon. While providing important novel evidence, it also has limitations future research may attempt to overcome.

When modeling our alternative hypothesis H1, for the sake of simplicity and in the absence of a theoretically or empirically justified theory, we opted for a simple additive method (see Eq 1). It is, however, possible that words' *Semantic Content* and PAP have an interactive effect on ratings of affective meaning. Similarly, the role of each phoneme in a word for contributing to the PAP might be differentially weighted depending on its position in the word (see Eq 2). Applying more sophisticated methods such as machine-learning-based regressors (e.g. [79,80]) might help integrate the large number of potentially influential factors into more complete and accurate models of the process of evaluative rating.

Also, a number of the acoustic features we used are measured as average frequencies, which precludes the use of dynamic sound features (e.g. spectral flex). Employing other methods based on dynamic changes of the sound signal might increase the accuracy of acoustic models predicting ratings of words' affective sound. A more sophisticated approach, for instance, might use the matrix of the spectrogram to quantitatively represent the sound envelope. Since the length of the audio signal (i.e. the length of words or pseudowords) differs for each item, the challenge of such an approach would be to find an appropriate method to classify the (pseudo)words' affective sound based on a series of independent variables, the number of which depends on physical signal length.

Alternatively, our acoustic analysis can be complemented by the use of phonetic categories (e.g. voiced/voiceless, obstruent/sonorant, etc.) to relate these categories to the effect of sound on words' affective meaning. In a simple phonetic approach, each phoneme in a word will represent a vector of phonetic features. Consequently, an entire word—comprising different phonemes—can be described as a concatenation of vectors of phonetic features, which can be used to calculate the contribution of any phonetic feature to the affective sound of words; in a similar fashion to our approach concerning PAVs. A practical approach concerning the use of phonetic features instead of acoustic variables would be the use of phonological cues defined as the proportion of consonants with particular manner and place features, and the average height and position of vowels (as provided in [81]). These cues can be used in the same way as our acoustic variables to identify phonological features underlying the PAPs. An advantage of this method would be the simple classification of the phonological construction of a word and its contribution to the sound to meaning relation. Our initial investigation has shown that such phonological cues can account for a significant portion of variance in the PAPs (25% for arousal, and 15% for valence), with the proportion of voiced consonants, and the average of vowel roundedness in a word being the most important predictors for both models of arousal and valence.

Another approach for measuring the affective sound of words can make use of the insights of sonority theory [82,83]. For this, each word can be assigned a sonority score which may also systematically contribute to affective (and aesthetic) ratings, as supported by recent findings concerning ratings of the aptness of metaphors and the beauty of words [79,84].

Practical applications

Our findings on the effects of implicit sound on affective meaning, and specifically our acoustic model for measuring the affective sound of words effectively suggest a method for constructing words and pseudowords associated with specific affects (positive/negative, arousing/calming) or emotions (e.g., fear, disgust), which can have broad applications in various contexts from marketing and advertising to art and literature. For instance, in the field of product and brand naming, previous work has shown that the sound of a product's name can in general set and modify consumer expectations about the likely attributes of the products [5,85] and that names with negative sounds were least preferred regardless of product category [86]. Here, our method for assessing the affective sound of words based on its acoustic features could provide a substantial improvement to previous work, which was usually based on the manipulation of a limited group of sounds (e.g. front vs. back vowels). Likewise, in artistic contexts, such as film, literature, and in particular, poetry, our method could be applied to evoke and verify particular emotional effects by use of words that possess specific implicit affective sounds.

Poetry is probably the best example of a sound meaning interaction: while it is inherently concerned with the expression and elicitation of emotions [5,56,87,88], it is deeply rooted at the perceptual level in the domain of sound [5,89–91]. Indeed, poetry has always artfully deployed sound patterns to shape order, to create a new layer of meaning, and to emphasize the affective meaning in a text. With the present study we provide a complementary method to previous attempts for analyzing poetic texts at the sublexical level [55,56,90–94], and for further examination of the influence of sound structure on affective and aesthetic reactions to verbal material intended to elicit a certain emotional impact in readers, such as advertisements, political speeches or manifests.

Conclusion

The present studies provide novel results on the contribution of the implicit sound of a word to its affective meaning. Our findings have the potential to shed new light on various unanswered questions regarding the evolution, organization, and processing of human language by drawing attention to the role of affect as well as by substantiating the psychological reality of iconicity in everyday language. These new insights may pave the way for further cross-linguistic investigations, as well as the detailed study of the neural substrates underlying the effect of phonology and sound-meaning interaction in language use; a phenomenon creatively exploited particularly by Poe and other poets throughout history.

Supporting information

S1 Database. Database_AM_AS_PAP.
(XLSX)

S1 Table.
(DOCX)

S2 Table.
(DOCX)

S3 Table.
(DOCX)

S4 Table.
(DOCX)

S5 Table.
(DOCX)

Author Contributions

Conceptualization: Arash Aryani, Markus Conrad, Arthur Jacobs.

Formal analysis: Arash Aryani, David Schmidtke, Arthur Jacobs.

Funding acquisition: Arthur Jacobs.

Investigation: Arash Aryani, David Schmidtke.

Methodology: Arash Aryani, Markus Conrad, Arthur Jacobs.

Project administration: Arash Aryani, Arthur Jacobs.

Resources: Arash Aryani, Arthur Jacobs.

Software: Arash Aryani.

Supervision: Arthur Jacobs.

Validation: Arash Aryani, Markus Conrad, David Schmidtke, Arthur Jacobs.

Visualization: Arash Aryani.

Writing – original draft: Arash Aryani.

Writing – review & editing: Arash Aryani, Markus Conrad, Arthur Jacobs.

References

1. De Saussure F. Course in general linguistics. Columbia University Press; 2011.
2. Poe EA. The bells. Philadelphia: Porter & Coates; 1881.
3. Jakobson R. Closing statement: Linguistics and poetics. *Style Lang.* 1960; 350–377.
4. Tsur R. What makes sound patterns expressive the poetic mode of speech perception. Duke University Press; 1992.
5. Schrott R., & Jacobs AM. Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren. Hanser; 2011.
6. Bowers JS, Pleydell-Pearce CW. Swearing, euphemisms, and linguistic relativity. *PLoS One.* 2011; 6. <https://doi.org/10.1371/journal.pone.0022341> PMID: 21799832
7. Carlin G. Doin'it Again. MPI Media Group; 1990.
8. Jakobson R. Lectures on sound and meaning. Cambridge, MA: MIT Press; 1937.
9. Sapir E. The status of linguistics as a science. *Language (Baltim).* JSTOR; 1929; 207–214.
10. Kohler W. Gestalt Psychology (1929). New York, NY Liveright. 1947;
11. Schmidtke DS, Conrad M, Jacobs AM. Phonological iconicity. *Front Psychol.* 2014; 5: 1–6. <https://doi.org/10.3389/fpsyg.2014.00001>
12. Dingemanse M, Blasi DE, Lupyan G, Christiansen MH, Monaghan P. Arbitrariness, Iconicity, and Systematicity in Language. *Trends Cogn Sci.* Elsevier Ltd; 2015; 19: 603–615. <https://doi.org/10.1016/j.tics.2015.07.013> PMID: 26412098
13. Lockwood G, Dingemanse M. Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Front Psychol.* 2015; 6: 1–14. <https://doi.org/10.3389/fpsyg.2015.00001>
14. Blasi DE, Wichmann S, Hammarström H, Stadler PF, Christiansen MH. Sound–meaning association biases evidenced across thousands of languages. *Proc Natl Acad Sci.* 2016; 113: 10818–10823. <https://doi.org/10.1073/pnas.1605782113> PMID: 27621455
15. Vigliocco G, Kita S. Language-specific properties of the lexicon: Implications for learning and processing. *Lang Cogn Process.* 2006; 21: 790–816. <https://doi.org/10.1080/016909600824070>

16. Perniss P, Vigliocco G. The bridge of iconicity: from a world of experience to the experience of language. *Philos Trans R Soc B Biol Sci.* 2014; 369: 20140179–20140179. <https://doi.org/10.1098/rstb.2014.0179>
17. Roberts G, Lewandowski J, Galantucci B. How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition.* Elsevier B.V.; 2015; 141: 52–66. <https://doi.org/10.1016/j.cognition.2015.04.001> PMID: 25919085
18. Imai M, Kita S. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos Trans R Soc B Biol Sci.* 2014; 369: 20130298–20130298. <https://doi.org/10.1098/rstb.2013.0298> PMID: 25092666
19. Monaghan P, Shillcock RC, Christiansen MH, Kirby S. How arbitrary is language? *Philos Trans R Soc B Biol Sci.* 2014; 369: 20130299–20130299. <https://doi.org/10.1098/rstb.2013.0299> PMID: 25092667
20. Thompson PD, Estes Z. Sound symbolic naming of novel objects is a graded function. *Q J Exp Psychol.* 2011; 64: 2392–2404. <https://doi.org/10.1080/17470218.2011.605898> PMID: 21895561
21. Ohala JJ. Ethological theory and the expression of emotion in the voice. *Proceeding Fourth Int Conf Spok Lang Process ICSLP '96.* 1996; 3: 1812–1815. <https://doi.org/10.1109/ICSLP.1996.607982>
22. Christiansen MH, Monaghan P. Division of Labor in Vocabulary Structure: Insights From Corpus Analyses. *Top Cogn Sci.* 2016; 8: 610–624. <https://doi.org/10.1111/tops.12164> PMID: 26399384
23. Farmer TA, Christiansen MH, Monaghan P. Phonological typicality influences on-line sentence comprehension. *Proc Natl Acad Sci.* 2006; 103: 12203–12208. <https://doi.org/10.1073/pnas.0602173103> PMID: 16882728
24. Bergen BK. The psychological reality of phonaesthemes. *Language (Baltim).* Linguistic Society of America; 2004; 80: 290–311.
25. Vigliocco G, Meteyard L, Andrews M, Kousta S. Toward a theory of semantic representation. *Lang Cogn.* 2009; 1: 219–247. <https://doi.org/10.1515/LANGCOG.2009.011>
26. Kousta ST, Vinson DP, Vigliocco G. Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition.* Elsevier B.V.; 2009; 112: 473–481. <https://doi.org/10.1016/j.cognition.2009.06.007> PMID: 19591976
27. Kousta ST, Vigliocco G, Vinson DP, Andrews M, Del Campo E. The Representation of Abstract Words: Why Emotion Matters. *J Exp Psychol Gen.* 2011; 140: 14–34. <https://doi.org/10.1037/a0021446> PMID: 21171803
28. Jacobs AM, Hofmann MJ, Kinder A. On elementary affective decisions: To like or not to like, that is the question. *Front Psychol.* 2016; 7: 1–17. <https://doi.org/10.3389/fpsyg.2016.00001>
29. Schmidtke DS, Aryani A. Bridges from affect to language Comment on “The quartet theory of human emotions: An integrative and neurofunctional model” by Koelsch S. et al. *Phys Life Rev.* Elsevier; 2015; 13: 83–84. <https://doi.org/10.1016/j.plrev.2015.04.024> PMID: 25912769
30. Osgood CE. The nature and measurement of meaning. *Psychol Bull.* American Psychological Association; 1952; 49: 197. PMID: 14930159
31. Mehrabian A. Basic dimensions for a general psychological theory implications for personality, social, environmental, and developmental studies. 1980;
32. Darwin C. *The descent of man and selection in relation to sex.* Murray; 1888.
33. Wundt WM. *Völkerpsychologie; eine untersuchung der entwicklungsgesetze von sprache, mythus und sitte: Bd. Die Kunst.* W. Engelmann; 1908.
34. Panksepp J. Emotional causes and consequences of social-affective vocalization. *Handbook of Behavioral Neuroscience.* Elsevier; 2010. pp. 201–208.
35. Ma W, Thompson WF. Human emotions track changes in the acoustic environment. *Proc Natl Acad Sci.* 2015; 112: 14563–14568. <https://doi.org/10.1073/pnas.1515087112> PMID: 26553987
36. Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex.* Elsevier Srl; 2012; 48: 788–804. <https://doi.org/10.1016/j.cortex.2010.11.002> PMID: 21163473
37. Gallese V, Lakoff G. The brain’s concepts: The role of the sensory-motor system in conceptual knowledge. *Cogn Neuropsychol.* Taylor & Francis; 2005; 22: 455–479. <https://doi.org/10.1080/02643290442000310> PMID: 21038261
38. Glenberg AM. Embodiment as a unifying perspective for psychology. *Wiley Interdiscip Rev Cogn Sci.* Wiley Online Library; 2010; 1: 586–596. <https://doi.org/10.1002/wcs.55> PMID: 26271505
39. Bühler K. *Sprachtheorie.* Fischer; 1934;
40. Jacobs AM, Võ MLH, Briesemeister BB, Conrad M, Hofmann MJ, Kuchinke L, et al. 10 years of BAWL-ing into affective and aesthetic processes in reading: What are the echoes? *Front Psychol.* 2015; 6: 1–15. <https://doi.org/10.3389/fpsyg.2015.00001>

41. Glenberg AM, Webster BJ, Mouilso E, Havas D, Lindeman LM. Gender, emotion, and the embodiment of language comprehension. *Emot Rev.* 2009; 1: 151–161. <https://doi.org/10.1177/1754073908100440>
42. Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, et al. The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nat Neurosci.* 2005; 8: 145–146. <https://doi.org/10.1038/nn1392> PMID: 15665880
43. Frühholz S, Trost W, Kotz SA. The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neurosci Biobehav Rev.* 2016; 68: 1–15. <https://doi.org/10.1016/j.neubiorev.2016.05.006>
44. Schirmer A, Kotz SA. Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci.* 2006; 10: 24–30. <https://doi.org/10.1016/j.tics.2005.11.009> PMID: 16321562
45. Brück C, Kreifelts B, Wildgruber D. Emotional voices in context: A neurobiological model of multimodal affective information processing. *Phys Life Rev. Elsevier B.V.*; 2011; 8: 383–403. <https://doi.org/10.1016/j.plrev.2011.10.002> PMID: 22035772
46. Brück C, Kreifelts B, Gößling-Arnold C, Wertheimer J, Wildgruber D. “Inner voices”: The cerebral representation of emotional voice cues described in literary texts. *Soc Cogn Affect Neurosci.* 2014; 9: 1819–1827. <https://doi.org/10.1093/scan/nst180> PMID: 24396008
47. Perrone-Bertolotti M, Kujala J, Vidal JR, Hamame CM, Ossandon T, Bertrand O, et al. How Silent Is Silent Reading? Intracerebral Evidence for Top-Down Activation of Temporal Voice Areas during Reading. *J Neurosci.* 2012; 32: 17554–17562. <https://doi.org/10.1523/JNEUROSCI.2982-12.2012> PMID: 23223279
48. Ziegler JC, Jacobs AM. Phonological information provides early sources of constraint in the processing of letter strings. *J Mem Lang. Elsevier*; 1995; 34: 567–593.
49. Braun M, Hutzler F, Ziegler JC, Dambacher M, Jacobs AM. Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Hum Brain Mapp. Wiley Online Library*; 2009; 30: 1977–1989. <https://doi.org/10.1002/hbm.20643> PMID: 18726911
50. Breen M. Empirical investigations of the role of implicit prosody in sentence processing. *Linguist Lang Compass.* 2014; 8: 37–50. <https://doi.org/10.1111/lnc3.12061>
51. Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic feature encoding in human superior temporal gyrus. *Science (80-)*. 2014; 343: 1006–1010. <https://doi.org/10.1126/science.1245994> PMID: 24482117
52. Ohala JJ. The frequency codes underlies the sound symbolic use of voice pitch.pdf. *Sound symbolism.* 1994. pp. 325–347.
53. Ullrich S, Kotz SA, Schmidtke DS, Aryani A, Conrad M. Phonological iconicity electrifies: An ERP study on affective sound-to-meaning correspondences in German. *Front Psychol.* 2016; 7. <https://doi.org/10.3389/fpsyg.2016.01200> PMID: 27588008
54. Myers-Schulz B, Pujara M, Wolf RC, Koenigs M. Inherent emotional quality of human speech sounds. *Cogn Emot.* 2013; 27: 1105–1113. <https://doi.org/10.1080/02699931.2012.754739> PMID: 23286242
55. Aryani A, Jacobs AM, Conrad M. Extracting salient sublexical units from written texts: “Emophon,” a corpus-based approach to phonological iconicity. *Front Psychol.* 2013; 4: 1–15. <https://doi.org/10.3389/fpsyg.2013.00001>
56. Aryani A, Kraxenberger M, Ullrich S, Jacobs AM, Conrad M. Measuring the Basic Affective Tone of Poems via Phonological Saliency and Iconicity. *Psychol Aesthet, Creat, Arts.* 2016; Vol 10(2);: 191–204. <http://dx.doi.org/10.1037/aca0000033>
57. Vö MLH, Conrad M, Kuchinke L, Urton K, Hofmann MJ, Jacobs AM. The Berlin Affective Word List Reloaded (BAWL-R). *Behav Res Methods.* 2009; 41: 534–538. <https://doi.org/10.3758/BRM.41.2.534> PMID: 19363195
58. Scherer KR. Vocal communication of emotion: A review of research paradigms. *Speech Commun.* 2003; 40: 227–256. [https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5)
59. Sauter DA, Eisner F, Calder AJ, Scott SK. Perceptual cues in nonverbal vocal expressions of emotion. *Q J Exp Psychol.* 2010; 63: 2251–2272. <https://doi.org/10.1080/17470211003721642> PMID: 20437296
60. Juslin PN, Laukka P. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychol Bull.* 2003; 129: 770–814. <https://doi.org/10.1037/0033-2909.129.5.770> PMID: 12956543
61. Duddington J. eSpeak 1.36. See <http://espeak.sourceforge.net>. 2008;
62. Dutoit T, Pagel V, Pierret N, Bataille F, der Vrecken O. The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Spoken Language, 1996 ICSLP 96 Proceedings, Fourth International Conference on.* 1996. pp. 1393–1396.

63. van Heuven VJ, Pols LC. Analysis and synthesis of speech: strategic research towards high-quality text-to-speech generation. Walter de Gruyter; 1993.
64. van Bezooijen R, van Heuven VJ. Assessment of synthesis systems. Gibbon D; Moore R; Winski R (ed), *Handb Stand Resour Spok Lang Syst*. Berlin/New York: Mouton de Gruyter; 1997; 481–563.
65. Boersma P, Weenik D. PRAAT: a system for doing phonetics by computer. Report of the Institute of Phonetic Sciences of the University of Amsterdam. Amsterdam Univ Amsterdam. 1996;
66. Bradley MM, Lang PJ. Affective norms for English words (ANEW): Instruction manual and affective ratings. 1999.
67. Keuleers E, Brysbaert M. Wuggy: A multilingual pseudoword generator. *Behav Res Methods*. 2010; 42: 627–633. <https://doi.org/10.3758/BRM.42.3.627> PMID: 20805584
68. Baayen RH, Piepenbrock R, van H R. The $\{\text{\$CELEX}\}$ lexical data base on $\{\text{\$CD-ROM}\}$. Linguistic Data Consortium; 1993;
69. Yarkoni T, Balota D, Yap M. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychon Bull Rev*. Springer; 2008; 15: 971–979. <https://doi.org/10.3758/PBR.15.5.971> PMID: 18926991
70. Leiner DJ. SoSci survey (version 2.4. 00-i)[computer software]. Accessed Febr. 2014; 20: 2015.
71. Bachorowski JA, Owren MJ. Vocal expression of emotion: Acoustic Properties of Speech Are Associated With Emotional Intensity and Context. *Psychol Sci*. 1995; 6: 219–224. <https://doi.org/10.1111/j.1467-9280.1995.tb00596.x>
72. Nyklíček I, Thayer JF, Van Doornen LJP. Cardiorespiratory differentiation of musically-induced emotions. *J Psychophysiol*. Hogrefe & Huber Publishers; 1997;
73. Boiten FA. The effects of emotional behaviour on components of the respiratory cycle. *Biol Psychol*. Elsevier; 1998; 49: 29–51. PMID: 9792483
74. Morton ES. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am Nat*. University of Chicago Press; 1977; 111: 855–869.
75. Bachorowski J-A. Vocal expression and perception of emotion. *Curr Dir Psychol Sci*. SAGE Publications Sage CA: Los Angeles, CA; 1999; 8: 53–57.
76. Kuhlmann M, Hofmann MJ, Briesemeister BB, Jacobs AM. Mixing positive and negative valence: Affective-semantic integration of bivalent words. *Sci Rep*. Nature Publishing Group; 2016; 6: 1–7. <https://doi.org/10.1038/s41598-016-0001-8>
77. Briesemeister BB, Kuchinke L, Jacobs AM. Emotion word recognition: Discrete information effects first, continuous later? *Brain Res*. Elsevier; 2014; 1564: 62–71. <https://doi.org/10.1016/j.brainres.2014.03.045> PMID: 24713350
78. Briesemeister BB, Kuchinke L, Jacobs AM, Braun M. Emotions in reading: Dissociation of happiness and positivity. *Cogn Affect Behav Neurosci*. Springer; 2015; 15: 287–298. <https://doi.org/10.3758/s13415-014-0327-2> PMID: 25398299
79. Jacobs A. M., & Kinder A. What makes a metaphor literary? Answers from two computational studies. *Metaphor Symb*. 2018;
80. Jacobs AM, Schuster S, Xue S, Lüdtkke J. What's in the brain that ink may character. . . . *Sci Study Lit*. John Benjamins Publishing Company; 2017; 7: 4–51.
81. Monaghan P, Christiansen MH, Chater N. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cogn Psychol*. 2007; 55: 259–305. <https://doi.org/10.1016/j.cogpsych.2006.12.001> PMID: 17291481
82. Clements GN. The role of the sonority cycle in core syllabification. *Pap Lab Phonol*. 1990; 1: 283–333.
83. Stenken P, Bastiaanse R, Huber W, Jacobs AM. Syllable structure and sonority in language inventory and aphasic neologisms. *Brain Lang*. 2005; 95: 280–292. <https://doi.org/10.1016/j.bandl.2005.01.013> PMID: 16246736
84. Jacobs AM. Quantifying the Beauty of Words: A Neurocognitive Poetics Perspective. *Front Hum Neurosci*. 2017; 11: 1–7. <https://doi.org/10.3389/fnhum.2017.00001>
85. Spence C. Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *J Consum Psychol*. Society for Consumer Psychology; 2012; 22: 37–54. <https://doi.org/10.1016/j.jcps.2011.09.004>
86. Baxter S, Lowrey TM. Phonetic symbolism and children's brand name preferences. *J Consum Mark*. 2011; 28: 516–523. <https://doi.org/10.1108/07363761111181509>
87. Lüdtkke J, Meyer-Sickendieck B, Jacobs AM. Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychol Aesthetics, Creat Arts*. 2014; 8: 363–377. <https://doi.org/10.1037/a0036826>

88. Jacobs AM. Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Front Hum Neurosci*. 2015; 9: 1–22. <https://doi.org/10.3389/fnhum.2015.00001>
89. Jacobs AM, Lüdtke J, Aryani A, Meyer-Sickendieck B, Conrad M. Mood-empathic and aesthetic responses in poetry reception. *Sci Study Lit*. John Benjamins Publishing Company; 2016; 6: 87–130.
90. Ullrich S, Aryani A, Kraxenberger M, Jacobs AM, Conrad M. On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic texts—a case study using 57 Poems of H. M. Enzensberger. *Front Psychol*. 2017; 7: 1–19. <https://doi.org/10.3389/fpsyg.2016.02073>
91. Kraxenberger M, Menninghaus W. Mimological reveries? Disconfirming the hypothesis of phono-emotional iconicity in poetry. *Front Psychol*. 2016; 7: 1–9. <https://doi.org/10.3389/fpsyg.2016.00001>
92. Auracher J, Albers S, Zhai Y, Gareeva G, Stavniychuk T. P is for happiness, N Is for sadness: Universals in sound iconicity to detect emotions in poetry. *Discourse Process*. 2011; 48: 1–25. <https://doi.org/10.1080/01638531003674894>
93. Whissell C. Phonosymbolism and the emotional nature of sounds: evidence of the preferential use of particular phonemes in texts of differing emotional tone. *Percept Mot Skills*. SAGE Publications Sage CA: Los Angeles, CA; 1999; 89: 19–48. <https://doi.org/10.2466/pms.1999.89.1.19> PMID: 10544399
94. Whissell C. Phonoemotional profiling: a description of the emotional flavour of English texts on the basis of the phonemes employed in them. *Percept Mot Skills*. SAGE Publications Sage CA: Los Angeles, CA; 2000; 91: 617–648. <https://doi.org/10.2466/pms.2000.91.2.617> PMID: 11065326