

TOWARDS EMOTION RECOGNITION IN TEXTS – A SOUND-SYMBOLIC EXPERIMENT

Dr. Velina Slavova, New Bulgarian University, Department of Computer Science, Bulgaria
E-mail: vslavova@nbu.bg

ARTICLE INFO

Original Research

Received: May, 30.2019.

Revised: July, 05.2019.

Accepted: July, 16.2019.

doi:[10.5937/IJCRSEE1902041S](https://doi.org/10.5937/IJCRSEE1902041S)

UDK

159.942.075-057.875:159.946.4(497.2)

Keywords:

Emotional sound-symbolism,

Text emotion recognition,

Emotion and cognition.

ABSTRACT

The purpose of this study is to investigate the relationship between the phonetic content of prose texts in English and the emotion that the texts inspire, namely - the effect of vowel-consonant bi-phones on subjects' evaluation of positive or negative emotional valence when reading. The methodology is based on data from an experiment where the participants, native speakers of three different languages, evaluated the valence invoked in them by one-page texts from English books. The sub-lexical level of the texts was obtained using phonetic transcriptions of the words and their further decomposition into vowel-consonant bi-phones. The statistical investigation relies on density-measures of the investigated bi-phones over each text as a whole. The result shows that there exists a correlation between the obtained sub-lexical representation and the valence perceived by the readers. Concerning the type of the consonants in the bi-phones (abrupt or sonorant), the influence of the abrupt bi-phones is stronger. However, sub-sets of both types of bi-phones showed relatedness with the emotional valence conveyed by the texts. In conclusion, the speech, expressed in written form, is laden with emotional valence even when the words' lexicological meaning is not taken into consideration and the words are apprehended as mere phonetic constructs. This prompts hypothesizing that words' semantics itself is partly underpinned by some mental emotion-related level of conceptualization, influenced by sounds. For practical purposes, the result suggests that based on the syllabic content of a text it should be possible to predict the valence that the text would inspire in its readers.

© 2019 IJCRSEE. All rights reserved.

1. INTRODUCTION

Several studies have shown that speakers from different cultures detect similar emotional content solely with reference to speech flow even when the speech consists of pseudo-words (e.g. [Scherer, 2000](#); [Scherer et al., 2001](#)). In the present study, focusing on the relationship between emotions, language, and speech, I have assumed from such evidence that several aspects of the speech, including its phonetic content, display some basic, species-typical pattern that involves features related to emotion transfer. The line of reasoning followed herein questions the principle of arbitrary

ness proposed in the theory of linguistic signs ([De Saussure, 1916/1983](#)).

Defining the essential attributes, functions and substrates of emotion has shown itself to be a particularly elusive problem, and overall consensus thereto remains lacking ([Kleinginna and Kleinginna, 1981](#); [Russell, 2003](#); [Hamann, 2012](#) for overviews). There exist two main models of emotions used in machine emotion recognition – a discrete model of emotion categories and a continuous model of emotion dimensions (for a recent overview see [Burton, 2015](#)). The set of Emotion categories (such as anger, disgust, fear, happiness, sadness, and surprise), as proposed in the works of [Ekman, P. \(1999\)](#) and further enlarged, has become accepted by the greater scientific community and as a foundation for the development of contemporary emotion recognition systems. The other model of emotion portrays emotional phenomena according to their discernible attributes, considered as emotion-dimensions. Following this model (suggested by [Russell and Mehrabian, 1977](#)) emotions are usually described with three di-

Corresponding Author

Dr. Velina Slavova, New Bulgarian University,

Department of Computer Science, Bulgaria

E-mail: vslavova@nbu.bg



This work is licensed under a [Creative Commons Attribution - NonCommercial - NoDerivs 4.0](#). The article is published with Open Access at www.ijcrsee.com

mensions - valence, arousal and dominance (VAD). Valence indicates whether emotions are pleasant (positive) or unpleasant (negative), arousal - the degree to which they are exciting or calming and dominance a rating of one's own status in relation to an emotion-causing occurrence. During the past few years, the VAD model has become intensively used in systems for emotion recognition as well as in the domain of sentiment analysis, prompting various efforts to bind dimensional models to emotional categories (e.g. [Buechel and Hahn, 2017](#)). However, the theoretical aspects regarding the brain mechanisms underlying the described formal models and the relation between them stay controversial.

In recent years, brain-imaging techniques have allowed the identifying of clear neural signatures that correspond to basic emotions. These patterns are involved deeply in the brain's typical multimodal fashion, exhibiting specific activations within a distributed network of cortical and subcortical areas (e.g., [Saarimäki et al., 2015](#)). In addition, questions regarding emotion dimensions and their correlates in terms of brain-functioning have been investigated. For example, [Maddock et al., \(2003\)](#) report fMRI evidence from a valence decision task showing the existence of specific brain regions activated only by unpleasant words and regions activated only by pleasant words. A recent study of brain-impaired subjects showed that both emotional valence and basic emotions are related to semantic memory, including for stimuli based on speech prosody ([Macoir et al., 2019](#)). Such results suggest that emotional valence is related to some semantic level of functioning of the human brain.

A matter of direct relevance to this study concerns emotional processing in relation to word recognition in reading. On the combined evidence of fMRI ([Marinkovic et al. 2003](#)) and event-related potential time-course studies (cf. [Abbassi, Kahlaoui et al. 2011](#), for a comprehensive review) it is known that, following initial recognition of visual features of words in the visual cortex, the resulting orthographic information is transmitted to auditory cortex and attributed its corresponding phonetic content. ([Van Orden, Johnston and Hale, 1988](#)). Recognition of lexical meaning proceeds largely in accord with processing of presented spoken word-sequences, that is – as perceived speech sounds.

There is mounting evidence that there exist non-arbitrary sound-symbolic patterns in language. One of the best-known examples

of systematic sound-symbolism and the first to be described ([Köhler, 1929](#)) is the correspondence of words predominantly consisting of abrupt consonants (e.g. t, k) with angular-shaped objects and words featuring sonorant consonants (e.g., m, l) with curvaceous objects. Subsequent research related to emotions ([Fantz and Miranda, 1975](#); [Leder, Tinio and Bar, 2011](#)) has demonstrated an inborn human emotional preference for curvaceous shapes.

During the last decade a huge amount of studies in sound-symbolism concentrate on the words' iconicity, a phenomenon considered as related mainly with the association between the sounds and the words' lexical meaning (e.g., [Imai and Kita, 2014](#); [Perlman, Dale and Lupyan, 2015](#), [Edmiston et al., 2018](#), [Winter et al., 2017](#); [Jones and Vigliocco, 2017](#), [Sidhu and Pexman, 2018](#) and many other works). The results of these numerous investigations confirm the relations between sound and meaning and propose that the phonetic content of the words could arise based on mechanisms related to perception, sensation, repeated imitation and so forth. These results suggest that sound-symbolic mechanisms have a general species-specific character. Indeed, in their recent experimental study, [D'Anselmo and colleagues \(D'Anselmo et al., 2019\)](#) did not find significant differences between Italian and Polish participants in guessing successfully the correct meaning of words in unknown languages. The authors conclude that there exist sound symbolic patterns that are independent of the mother tongue of the listener.

The connection between emotions and sound symbolic features also started to gain attention and led to results showing, for example, that taste and smell words form an affectively loaded part of the English lexicon ([Winter, 2016](#)).

The study of phonological effects on emotion has centered predominantly on poetry. The reasoning related to poetry originates from the Russian Formalists where the sound in poetry was first examined in a systematic fashion (for descriptions see [Trotsky, 1957](#); [Jakobson, 1960](#); [Shklovsky, 1990](#); [Mandelker, 1983](#)). Subsequent research, focusing on phoneme-level effects on experienced emotional state, has examined diverse examples of English poetry ranging from Byron to Beatles' song lyrics (see [Whissell, 1999](#)). Recent emotion-focused studies of German poetry undertaken by [Aryani, Ullrich, and colleagues \(Aryani et al., 2013](#); [Aryani et al., 2016](#); [Ullrich et al, 2017](#)) have also demonstrated the existence of a relation between the phonetic content of

poetic texts and the emotion that they convey.

The investigations of more general emotion-effects implicated in sound-symbolism have revealed important findings. Kawahara and Shinohara (2012), for example, conducted experiments, demonstrating a tripartite trans-modal symbolic relationship between three domains of cognition (auditory sounds, visual shapes, and emotions). A recent study by Adelman and colleagues (Adelman et al., 2018), based on data-analysis of corpora representing five languages of Europe, revealed the existence of strong correlations between, for example, abrupt initial-position phonemes and highly negative, arousing words as well as between slow onset initial phonemes and positive words. These recent studies concentrate on the general effect of single phonemes.

The idea that syllables represent basic compositional elements of speech is not new in linguistics (see, e.g., Itō, 2018). The rationale of consonant-vowel speech compositional sequences is incorporated in the written form of several languages, for example, Arabic. A syllabic structure is detectable, too, in sign-languages. For example, a recent study by Gökgöz (2018) revealed that Turkish sign-language has a syllabic composition. The syllables-emotion relation is acknowledged and used in speech-emotion recognition (e.g. Origlia et al., 2014). The role of syllables as emotional indicators was investigated, too, in terms of micro-prosody conveyed by syllabic pitch-profiles (Brandt and Bennett, 2015).

The study proposed here investigates the relation between the syllabic content of prose texts in the English language and the conveyed emotional valence with the hypothesis that such a relation exists.

2. MATERIALS AND METHODS

2.1. Goal and approach

The goal of the presented work is to explore statistically the relation between the vowel-consonant syllabic content of prose texts and the emotional valence that the text inspires in the readers in order to provide some phonological and statistical details of such relation, if it exists.

The used approach examines only the phonetic content of the word-sequences, without regard to their lexicological meaning, speech prosody, and other features of language that are commonly considered related to communicating emotion.

The study proposed here was inspired

by the analysis of the tripartite trans-modal symbolic relationship by Kawahara and Shinohara (2012). Their study was based on pseudo-words composed by two consonant-vowel syllables, organized in two types of phonological stimuli: “abrupt” (“Stop condition” - e.g. [tadi]) and “sonorant” (“Sonorant condition” - e.g. [maji]). The obtained result, based on auditory input of such pseudo-words, confirmed that English speakers associate oral stop consonants with angular shapes and showed that oral stops are associated with emotion types that involve abrupt onsets (e.g., “shocked” and “surprised”) and, too, that angular shapes are associated with those types of emotions that involve abrupt onsets. The study has led to the same kind of conclusions concerning the sonorant condition.

In the approach proposed in the present study, I supposed that features as “oral stop” and “oral passage” are clearly detectable in biphone syllables of the type vowel-consonant. The statistical treatment of the experimental data performed here is based solely on this syllabic scheme. Two forms of vowel-consonant biphones - vowel-abrupt (e.g., /ot/) and vowel-sonorant (e.g., /um/) were deployed. On the presumption that the emotional dimension of valence has a neuropsychological basis, I undertook an exploration of the relation between the English language, represented at vowel-consonant sub-lexical level, and emotional valence.

2.2. Experiment and data

In testing the hypothesis, the aim was to extract from textual data some metadata that is necessary for statistically investigating the presence of a relatedness of the phonetic characteristics of texts and the valence inspired in the readers. This was done based on experimental data, where the experiment was designed for this specific purpose (Figure 1).

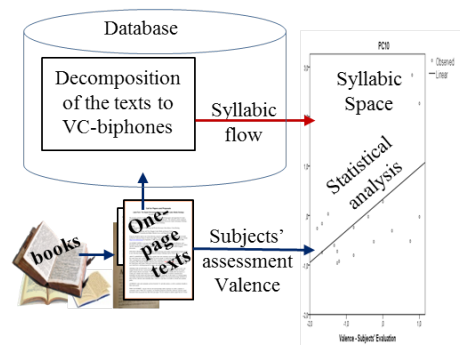


Figure 1. General scheme of the study

The experiment was organized and conducted within a student’s project at the New Bulgarian University. Twenty approximately one-page-long passages of text were selected, each evoking a certain scenario that would predictably inspire a given emotional valence. The passages (hereafter called *experimental texts*) were taken from works by Leigh Bardugo, Charles Dickens, Neil Gaiman, Robin Hobb, Derek Landy, Brandon Sanderson, and J. R. R. Tolkien.

The participants in our experiment were residents of different countries, were native speakers of different languages (Arabic=2, Bulgarian=10, English=5), and all were highly proficient English language users. Their task was to evaluate each text overall and not its parts.

We submitted four texts (that we judged as different with regards of valence) to each participant to read silently and evaluate the valence of each. All participants received in written form the following instruction:

“Many texts invoke certain emotions in the reader. For example, “A rose by any other name would smell as sweet” by Shakespeare invokes joy, positive emotion, while “As the light begins to intensify, so does my misery, and I wonder how it is possible to hurt so much when nothing is wrong” by Tabitha Suzuma – sadness, negative emotion. You will receive four texts in English. They are excerpts (a page long) from famous literary works, written by English natives. Evaluate what emotion is invoked in you by each of these texts by using this scale:

Very negative, Negative, Does not arouse emotion, Positive, Very positive “

The received written evaluations were assigned numeric marks from -2 (very negative) to +2 (very positive). The 20 texts were evaluated by the 17 participants, where each text was evaluated by at least 3 subjects (Max 6, Mean 4.7).

2.3. Phonetic representation and sub-lexical metadata

The textual data was stored in a database (see Figure 3) and further treated in order to extract its phonological characteristics as a stream of syllables. To perform the extraction of syllabic metadata from the raw text, 6 main steps were accomplished, as illustrated in Figure 2.

The sentences were first decomposed into word-forms. This step identified 10,692 word occurrences in the experimental texts. Next, the words’ phonetic transcriptions (following the commonly adopted IPA standard of English phonetic transcription) were downloaded from the Internet using online dictionaries. A number of words (about 2000) have necessitated a manual search. The dictionary of the transcribed words obtained after this step contains 5,767 phonetically transcribed words. The vocabulary used in the experimental texts contains 2,505 different English word-forms where 1,989 (79% of the used vocabulary) were represented with their transcription.

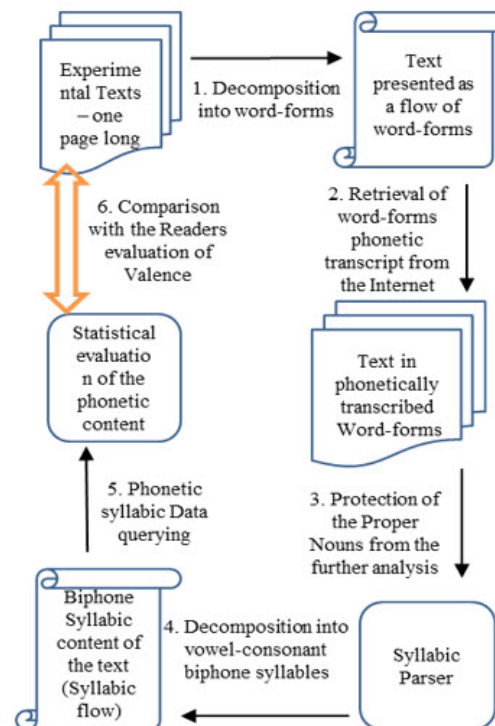


Figure 2. Stages of the treatment

Next, all word-occurrences which are proper nouns were retrieved and marked so as to not be included in the further analysis (see the sentence-example in Table 1). This step was necessary in order to exclude them from the emotion-related statistical picture, because

such words are not necessarily chosen by the writer and hence do not reflect her emotional state or intent. After this protective step, 153 occurrences of proper nouns were excluded. In total, 95% of the word occurrences in the experimental texts were subjected to further phonetic-decomposition and statistical treatment.

The next step was the syllabic decomposition of the transcripts. The approach proposed here is based on biphone syllables based on the entire set of phonemes in the English language. We took the set V of vowels (/æ/, /ɒ/, /ʌ/, /i:/, etc.) and the set C of consonants (/ŋ/, /d/, /k/, /f/, etc.) and composed the Cartesian product $V \times C$ containing all possible biphone syllables of one vowel ($v \in V$) as first element and one consonant ($c \in C$) as a second element. Further, the biphone syllables obtained this way are called here “VC-biphones”.

The obtained Cartesian product, containing all 528 possible combinations of vowels and consonants, was stored in a separate table (Table “Syllables” in Figure 3) and used to decompose the phonetic transcription of words into VC-biphones when parsing. Parsing of the transcribed words showed that 267 out of the 528 VC-biphones are used in the experimental texts overall. In this way, the words’ phonetic transcriptions were presented as a VC-biphone sequence for each word and, consequently, a VC-biphone sequence for each text (see the example in Table 1).

Further in this paper, I call a syllabic flow the phonological image of the experimental texts represented by means of VC-biphones, as shown in the example provided in Table 1. For the phonological analysis, with regards of the results reported by Kawahara and Shinohara (2012), the VC-biphones were divided into “abrupt” subset (/ɑ:ð/, /ɒdʒ/, /æb/, /aʊtʃ/, etc.) and “sonorant” subset (/ɑ:m/, /ɒn/, /ɔ:r/, /əʊl/, /əw/, etc.) following their consonant (voiced or unvoiced).

Table 1. Example of decomposition of the sentence “Sarene stepped off of the ship to discover that she was a widow.” into VC-biphones – a query result.

Word flow	po-si-tion	Phonetic transcript	Syllabic flow	se-quence
Sarene	1	@Sarene		
stepped	2	stept	ep	1
off	3	ɒf	ɒf	1
of	4	əv	əv	1
the	5	'ðə		
ship	6	ʃɪp	yp	1
to	7	'tu:		
discover	8	dɪskʌvər	ys	1
discover	8	dɪskʌvər	ʌv	2
discover	8	dɪskʌvər	ər	3
that	9	ðæt	æt	1
she	10	'ʃi:		
was	11	wɒz	ɒz	1
a	12	'eɪ		
widow		wɪdɒw	yd	1

It should be noted that the particularities of standard notation used in the transcripts led to naming the phoneme /ɪ/ (as per e.g. /ɪt/ in [ˈbenɪfɪt]) to /y/ because the upper case of the letter i used to denote the phoneme is undistinguishable from its lower case (as per, e.g., /ɪt/ in [rɪt]) by the used data-treatment product.

As the statistical treatment was based on results of counting queries, it was important to protect the calculations from erroneous data-fusion, and the data was organized in a relational database in 3th Normal Form with referential integrity (Figure 3). This means of structuring the data allowed performing the obligatory step of data-verification based on a comprehensible text-decomposition as shown in Table 1, and thereupon, to perform count-queries.

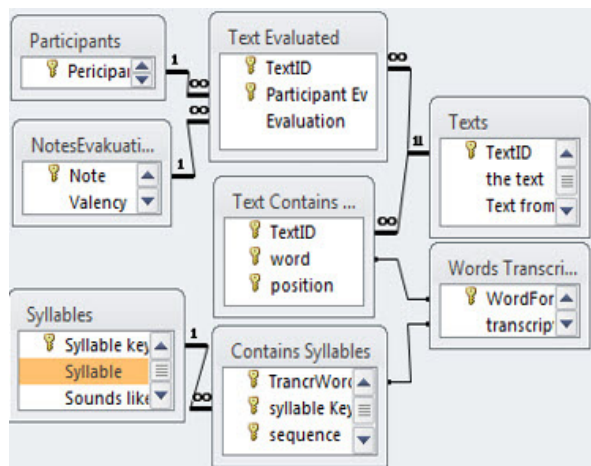


Figure 3. Data Base - organization of the data for the treatment

The analysis showed that the 20 experimental texts were containing on average 534 words (min. 207, max. 1016). The overall number of VC-biphones detected in the experimental texts is 9,275 where 4,624 are abrupt and nearly the same number – 4,651, are sonorant. The phonetic density of the examined feature seen as percentage of VC-biphones over the total number of words is quite low – 90%, that is – on average, one word contains less than one VC-biphone.

3. RESULTS

3.1. Correlation between emotional valence and the volume of abrupt syllables

To evaluate the content of each experimental text in terms of participation of abrupt and sonorant VC-biphones in the syllabic flow, the following measure for the Texts' Syllabic Charge (TSC) was applied:

$$(1) \quad TSC_{STj} = \frac{NSyll_{STj}}{NWord_j}$$

where the index ST denotes the type of the VC-biphones – abrupt or sonorant, NSyll_{STj} shows the number of VC-biphones of type ST that appear in the experimental text j, NWord_j is the number of transcribed words in the experimental text j (j = 1 to 20). The texts syllabic charge shows the extent of involvement of abrupt or of sonorant VC-biphones in the syllabic flow in a given experimental text.

This measure revealed that the used textual data cannot provide a reliable statistical picture for the texts' charge regarding the so-

norant VC-biphones and their correlation with valence (Pearson's $r = -0.06$, $p < 0.8$).

The text syllabic charge ensuing from abrupt VC-biphones led to much more reliable statistics, indicating relatedness between the volume of abrupt VC-biphones and Valence ($r = 0.416$, $p < 0.068$). This indicates that at least some subset of abrupt VC-biphones has influenced the readers' emotional judgment. The plot in Figure 4 illustrates this tendency.

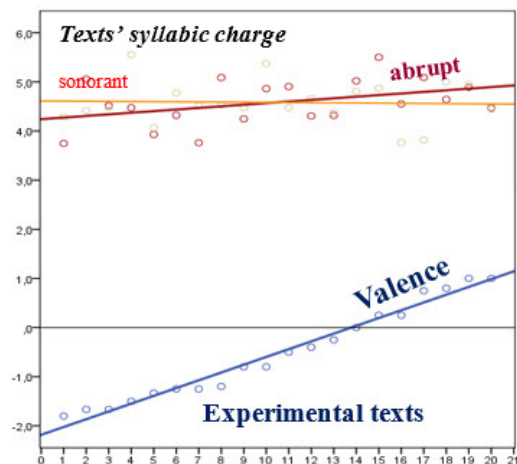


Figure 4. Texts' Valence and Syllabic Charge of abrupt and sonorant VC-biphones

The displayed result is not statistically unquestionable, but nevertheless indicates a particular tendency. This fact prompted two hypotheses: 1. The experimental texts are too short to contain a sufficient volume of VC-biphones; 2. The split of the VC-biphones into abrupt and sonorant sub-sets does not lead to a clear-cut statistical picture. The next step was to investigate the influence of each of the syllables separately, independently of their abruptness.

3.2. Connection between emotional valence and the set of VC-biphones

To evaluate the degree of involvement of each of the VC-biphones in the syllabic flow pertaining to each of the experimental texts, a Syllabic Ratio per Document (*RatSyll*) of each of the 267 biphones appearing in the texts was calculated using the equation:

$$(2) \quad RatSyll_{ij} = k \cdot \frac{NSyll_{ij}}{NWord_j \cdot NSyll_j}$$

where $NSyll_{ij}$ shows how many times the VC-biphone i ($i=1$ to 267) appears in the

experimental text j , $NWord_j$ is the number of transcribed words in the experimental text j ($j = 1$ to 20) and $NSyll_j$ is the number of VC-biphones in the syllabic flow of the experimental text j . As the Syllabic Ratios obtain very small values, for reasons related to the perceptibility of the metadata, they were multiplied by $k=10^6$. The Syllabic Ratio per Document shows the extent of involvement of a given VC-biphone in the syllabic flow of a particular experimental text.

The correlation analysis of the interdependency between the Valence-score and the Ratios of the VC-biphones showed that the investigated textual data provides a reliable statistical result (p -value < 0.09) for the 13 VC-biphones listed in Table 2. As it can be observed, only two of these VC-biphones are from the sonorant subset. Some of the investigated VC-biphones display a high and reliable correlation between them, because, in general, the words in the language have several compositional rules to respect.

Table 2. Pronounced Correlation of separate VC-biphones with the Valence score

VC-biphone	correlation with Valence	p-value	type
ʊt	0.615	0.0039	abrupt
ys	0.552	0.0116	abrupt
æv	0.489	0.0286	abrupt
yk	0.487	0.0293	abrupt
əs	0.487	0.0295	abrupt
ert	0.464	0.0392	abrupt
ydʒ	0.422	0.0636	abrupt
eɪp	0.414	0.0693	abrupt
u:t	0.405	0.0768	abrupt
ɜ:n	0.389	0.0900	Sonorant
u:l	-0.413	0.0702	Sonorant
i:θ	-0.473	0.0353	abrupt
eɪs	-0.500	0.0248	abrupt

As, due to the nature of the language system itself, the phonemes in the speech are correlated between them, in order to detect a comprehensive statistical picture, a dimension reduction using principal component analysis (PCA) was performed.

The 20 experimental texts were presented in a 267-dimensional statistical space in which each document is depicted by the Syllabic Ratios derived from its corresponding syllabic flow. The initial 267-dimensional

space was reduced (having eigenvalues > 1 , where 99% of the variance was extracted) to 18 principal components (PCs).

This reduced space expresses features which are discriminative for the 20 experimental texts and are derived from their syllabic flows. It should be noted that, due to the rather limited amount of textual data, some of the VC-biphones have been excluded from the analysis as they occur only a few times in the textual data and/or occur with zero variance in the set of experimental documents. The number VC-biphones which syllabic ratios were submitted to PCA is 174. The coordinates of the experimental texts were recalculated in the obtained 18-dimensional space, called hereafter a *PC Syllabic space*. Figure 5 shows the experimental texts presented on the first 3 PCs.

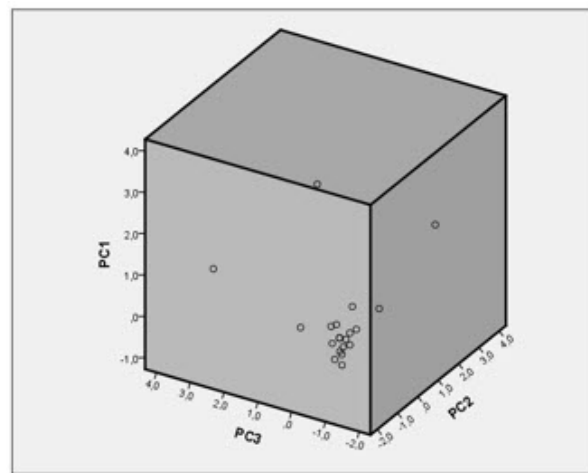


Figure 5. The 20 experimental texts, presented in the PC Syllabic space (the first 3 PCs).

The relationship between emotional valence and the phonetic features expressed by the PC Syllabic space were investigated in terms of correlation between the valence-scores of the 20 texts and the 20 coordinates of these texts on the 18 axes (the PCs) of the PC Syllabic space.

The correlation analysis revealed that one of the PCs displays a reliable and important correlation with valence. (PC#10: Pearson's $r = 0.607$, $p < 0.005$). The plot of the linear regression of this PC on Valence is shown in Figure 6. No other PCs displayed a reliable and important correlation with valence.

This result suggests that there exist a relationship between the VC-biphone content of a text and the emotional valence that the text inspires in the reader.

Next, a check was conducted to ascer-

tain whether the abrupt VC-biphones had some more important impact on the obtained valence-related PC (seen as expressing text-discriminative biphonic feature) as such an impact was suggested by the correlation of the texts syllabic charge reported in section 3.1.

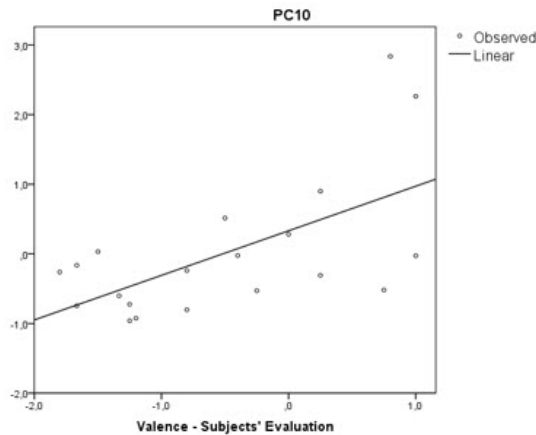


Figure 6. Interdependency between the valence-scores of the texts and their phonetic metadata.

3.3. Impact of the abrupt and sonorant VC-biphones

The further step was to investigate separately the impact of abrupt and sonorant VC-biphones on valence using the coefficients in the PCA transition matrix. The assumption was that the PC which shows a high correlation with the valence-score expresses some “summarized” and independent feature present in the syllabic flow which is important for inspiring valence.

The total number of VC-biphones that are projected on the valence-correlated PC is 174 where 118 are a subset of the abrupt VC-biphones and only 56 – a subset of the sonorant VC-biphones. The plot of the values of their corresponding transition coefficients is given in Figure 7. As it is seen, both types of VC-biphones are projected with both - positive and negative transition coefficients. It is also seen that the subset of abrupt VC-biphones has a stronger effect and that their effect is mostly in positive direction, confirming the result reported in section 3.1. The subset of sonorant VC-biphones has a smaller effect on valence, but, scientifically, this effect cannot be dismissed. The list of the first 10 “more negatively” and “more positively” projected VC-biphones of both types is provided in Table 3.

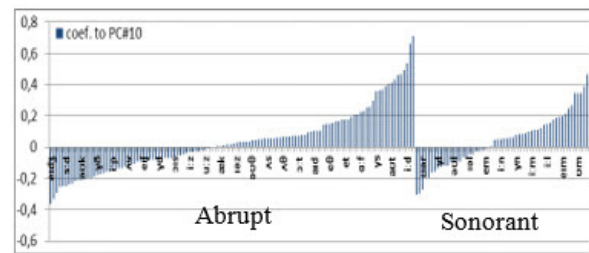


Figure 7. Plot of the influence of the VC-biphones on the valence-correlated PC, expressed by means of their coefficients in the component transition matrix.

It can be seen in Table 3 that the VC-biphones /eip/ and /yk/ (/Ik/), positively projected on the valence-correlated PC, are between the listed in Table 2 VC-biphones which, even in the small amount of textual data used in the experiment, displayed, each of them separately, a reliable positive correlation with valence.

As indicated by the extracted metadata, the vowels included in the VC-biphones do not represent a factor which seems related with the inspired valence. For example, as seen in Table 3, the vowel /e/ appears in the abrupt /eip/ and in the sonorant /eim/ VC-biphones, which have a positive influence and, the same vowel /e/ appears in abrupt /eidz/ and in the sonorant /eil/, which have a negative influence.

The same type of contradictory inclusion of vowels can be observed in Table 2 for the vowel /u:/ which is included in VC-biphones displaying both - negative and positive correlation with valence. Such observations suggest that an eventual detailed analysis has to take into consideration more specific phonological features of the vowels.

Table 3. The VC-biphones assigned with the ten top positive and negative coefficients by PCA.

Abrupt	coef.	Sonorant	coef.
ep	0.709	ɑ:n	0.467
ɒb	0.660	ɔ:l	0.390
i:d	0.535	ɜ:n	0.347
ɑ:ð	0.492	ʊm	0.345
ɒf	0.464	aʊn	0.343
yk	0.458	aɪl	0.267
əʊz	0.434	ər	0.249
aʊt	0.408	u:n	0.213
əb	0.397	eɪm	0.200
ʊd	0.392	ɪŋ	0.193
eɪdʒ	-0.364	ʌŋ	-0.305
ef	-0.333	ɑ:r	-0.297
əp	-0.291	ʊər	-0.272
ɑ:s	-0.256	ɒŋ	-0.196
ɪv	-0.251	ɒm	-0.194
ɜ:d	-0.249	æn	-0.163
i:tʃ	-0.237	eər	-0.159
ep	-0.236	ɪl	-0.133
ɪp	-0.219	ʊl	-0.126
i:θ	-0.213	eɪl	-0.120

4. DISCUSSION

4.1. Next steps to be performed

The analysis proposed here is based on a subset of syllables which density in the texts is quite low. The step to be performed using the same method is to include in the syllabic flow the consonant-vowel biphones and to investigate a greater amount of valence-evaluated textual data.

The selection of valence-relevant texts represents a concern because they have to be long enough in order to contain a representative subset of syllables and, at the same time, the content of each must be such as to evoke, in an overall homogenous way, a similar rating of valence in readers. This problem can be solved using existing corpora of emotionally evaluated texts combined with convenient strategies for assembling valence-homogenous data-sources with appropriate volume.

4.2. General discussion

From the result presented here, it is not possible to explain the principle by which the observed phonetically accomplished transmitting of emotion-encoding information enters and exerts its emotional effect on a text's lexical substance. To my knowledge, there is no ready-to-hand explanation of the phenomenon described here. The key questions awaiting scientific explanation are: 1. how did words come to incorporate these sound-patterns and 2. to what extent the observed patterns are language-dependent? The parts of such puzzle are far from being assembled. Comparative studies of more languages could identify some general language-independent features. Key to the puzzle is to understand the manner in which humans construct their semantic representations.

The result of this study prompts the hypothesis that concept's semantics itself is partly underpinned by some emotion-related level of representing phenomena, rooted in the long evolutionary development of animal species, while phenomenal iconicity and sound symbolism are instrumental for its overt, verbally framed expression as human speech.

5. CONCLUSION

The statistical result shows that there exists emotion-related information incorporated in the VC-biphone content of the English language. Thus, the speech is laden with emotional meaning even when the words' lexicological meaning is not taken into consideration and the words are apprehended as mere phonetic constructs. The study's general conclusion is that the syllabic composition of the words is not arbitrary from the standpoint of valence.

The result suggests that phonological characteristics prevailing within a syllabic flow should make it possible to predict the valence that a given text would inspire in its readers. In other words, the features of the syllabic flow could be useful for valence-classification of texts.

However, the statistical parameters of the proposed analysis suggest that the concrete details of the revealed interdependency can be correctly assessed using experimental data of larger volume.

ACKNOWLEDGEMENTS

The realization of several technical aspects related to the collected experimental data was organized by myself as instructor for the course “CSCB024” in programming and use of Internet resources - a regular practical course for graduate computer science students at the New Bulgarian University. The author is thankful to her colleague Dr. Filip Andonov and to the Ph.D. student Marwan Soula for their involvement in the text-processing task, and to the students: Lazar Dilov - for developing the syllabic parser and Shteriana Kostova – for selecting the experimental texts and organizing the access to them, as well as to all the participants in the experiment. Sincere gratitude to Richard Traub for his interest and his participation in the research exposed here, during the work on which he prepared and edited several parts of the presented text.

Conflict of interests

The author declares no conflict of interest.

REFERENCES

- Abbassi, E., Kahlaoui, K., Wilson, M. A., & Joannette, Y. (2011). Processing the emotions in words: The complementary contributions of the left and right hemispheres. *Cognitive, Affective, & Behavioral Neuroscience*, 11(3), 372-385. <https://doi.org/10.3758/s13415-011-0034-1>
- Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, 175, 122-130. <https://doi.org/10.1016/j.cognition.2018.02.007>
- Aryani, A., Conrad, M., & Jacobs, A. M. (2013). Extracting salient sublexical units from written texts: “Emophon,” a corpus-based approach to phonological iconicity. *Frontiers in psychology*, 4, 654. <https://doi.org/10.3389/fpsyg.2013.00654>
- Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, 10(2), 191. <https://doi.org/10.1037/aca0000033>
- Brandt, P. A., & Bennett, A. (2015). ‘It’s Five O’Clock’ – Microprosody and Enunciation. Available at SSRN 2566228. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2566228
- Buechel, S., & Hahn, U. (2017, April). EMOBANK: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 578-585). <https://doi.org/10.18653/v1/E17-2092>
- Burton, N. L. (2015). Heaven and hell: The psychology of the emotions. Kent, UK: Acheron Press.
- D’Anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing meaning from word sounds of unfamiliar languages: a cross-cultural sound symbolism study. *Frontiers in Psychology*, 10, 593. <https://doi.org/10.3389/fpsyg.2019.00593>
- De Saussure, F. (1916/1983). *Course in General Linguistics* (trans. Roy Harris). London: Duckworth.
- Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences*, 285(1874), 20172709. <https://doi.org/10.1098/rspb.2017.2709>
- Ekman, P. (1999). *Basic emotions*. Hand-book of cognition and emotion, 45-60. <https://doi.org/10.1002/0470013494.ch3>
- Fantz, R. L., & Miranda, S. B. (1975). Newborn Infant Attention to Form of Contour. *Child Development*, 46(1), 224. <https://doi.org/10.2307/1128853>
- Gökgöz, K. (2018). Syllables in TİD. *Dilbilim Araştırmaları Dergisi*, 29(1), 29-49. <https://doi.org/10.18492/dad.374893>
- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*, 16(9). <http://doi.org/10.1016/j.tics.2012.07.006>
- 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. <https://doi.org/10.1098/rstb.2013.0298>
- Itô, J. (2018). *Syllable theory in prosodic phonology*. Routledge. <https://doi.org/10.4324/9780429455124>
- Jakobson, R. (1960). *Linguistics and poetics*. In *Style in language* (pp. 350-377). MA: MIT Press.
- Jones, J. M., & Vigliocco, G. (2017). Iconicity in Word Learning: What Can We Learn from Cross-Situational Learning Experiments?. *Proceedings of the Annual Meeting of the Cognitive Society 2017*. Retrieved from: <https://pdfs.semanticscholar.org/ac61/7f1f53f8cc4df37b315edb4672df86294109.pdf>
- Kawahara, S., & Shinohara, K. (2012). A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (p. Vol 34, No 34). Retrieved from <https://pdfs.semanticscholar.org/3408/1ce795c89940e8fdd2a23b97a6aa834319dc.pdf>
- Kleinginna, P. R., & Kleinginna, A. M. (1981). A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. *Motivation and Emotion*, 5(4), 981. <https://doi.org/10.1007/BF00992553>
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Leder, H., Tinio, P. P., & Bar, M. (2011). Emotional valence modulates the preference for curved objects. *Perception*, 40(6), 649-655. <https://doi.org/10.1068/p6845>
- Macoir, J., Hudon, C., Tremblay M-P, Laforce R. Jr & Wilson M. A. (2019). The contribution of semantic memory to the recognition of basic emotions and emotional valence: Evidence from the semantic variant of primary progressive aphasia,

- Social Neuroscience*, <https://doi.org/10.1080/17470919.2019.1577295>
- Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human brain mapping*, 18(1), 30-41. <https://doi.org/10.1002/hbm.10075>
- Mandelker, A. (1983). Russian formalism and the objective analysis of sound in poetry. *Slav. East Eur. J.* 27, 327-338. <https://doi.org/10.2307/307860>
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3), 487-497. [https://doi.org/10.1016/S0896-6273\(03\)00197-1](https://doi.org/10.1016/S0896-6273(03)00197-1)
- Origlia, A., Cutugno, F., & Galatà, V. (2014). Continuous emotion recognition with phonetic syllables. *Speech Communication*, 57, 155-169. <https://doi.org/10.1016/j.specom.2013.09.012>
- Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, 2(8), 150152. <https://doi.org/10.1098/rsos.150152>
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145. <http://doi.org/10.1037/0033-295X.110.1.145>
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a Three-Factor Theory of Emotions. *Journal of Research in Personality*, 11, 273-294. [http://doi.org/10.1016/0092-6566\(77\)90037-X](http://doi.org/10.1016/0092-6566(77)90037-X)
- Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., ... & Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. *Cerebral cortex*, 26(6), 2563-2573. <https://doi.org/10.1093/cercor/bhv086>
- Scherer, K. R. (2000). A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. In *Sixth International Conference on Spoken Language Processing*. Retrieved from: https://www.isca-speech.org/archive/archive_papers/icslp_2000/i00_2379.pdf
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, 32(1), 76-92. <https://doi.org/10.1177/0022022101032001009>
- Shklovsky, V. (1990). *Theory of Prose*, transl. B. Sher. (Elmwood Park, IL: Dalkey Archive).
- Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic bulletin & review*, 25(5), 1619-1643. <https://doi.org/10.3758/s13423-017-1361-1>
- Trotsky, L. (1957). *Literature and Revolution*. New York, NY: Russell and Russell.
- Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., & Conrad, M. (2017). 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 HM Enzensberger. *Frontiers in psychology*, 7, 2073. <https://doi.org/10.3389/fpsyg.2016.02073>
- Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 371. <https://doi.org/10.1037//0278-7393.14.3.371>
- Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, 31(8), 975-988. <https://doi.org/10.1080/23273798.2016.1193619>
- Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?. *Interaction Studies*, 18(3), 443-464. <https://doi.org/10.1075/is.18.3.07win>
- Whissell, C. (1999). Phonosymbolism and the Emotional Nature of Sounds: Evidence of the Preferential Use of Particular Phonemes in Texts of Differing Emotional Tone. *Perceptual and Motor Skills*, 89(1), 19-48. <http://doi.org/10.2466/pms.1999.89.1.19>

