Universitat Pompeu Fabra Barcelona

Treball de fi de màster de Recerca

Examining Iconicity Information in Semantic and Phonetic Word Embeddings

Nom i Cognoms Jamie Wright

Màster: Lingüística Teòrica i Aplicada

Edició: 2022-202

Directors: Dr Thomas Brochhagen

Any de defensa: 2023

Col·lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

Abstract

This study explores the concept of iconicity, which refers to the relationship between sound and meaning in words. The Bouba-kiki effect, where participants consistently associate certain phonemes with round or spiky shapes, and ideophones, which are sensory words like *zigzag,* are examples of iconicity. In this study, iconicity ratings, based on native speaker judgments, are used to assess iconicity, while semantic and phonetic embeddings provide distributional and articulatory information about words. Linear regression models are employed to connect these elements, investigating the extent to which iconicity information is captured in English phonetic and semantic word embeddings. The models successfully predict iconicity ratings to a certain degree, suggesting the presence of iconicity information in these embeddings. Moreover, the models' predictions align with previous research on the phonetic and semantic dimensions of iconicity. This research shines light on the relationship between iconicity and word embeddings, and contributes to our knowledge on the cognitive and linguistic mechanisms underlying the phenomenon of iconicity.

Keywords: iconicity, non-arbitrariness, sound symbolism, cross-modal correspondence

Table of Contents

Introduction

The relationship between the linguistic sign and its meaning is one that has been the subject of debate throughout the history of linguistic inquiry. The conventional notion in linguistics of the 'arbitrariness of the sign', that there is no connection between the phonetic form of a word and its meaning, was perhaps most famously stated in Saussure's (1972: 67-68) 'first principle', where he argues that there is no inherent connection "between the idea 'sister' and the French sequence of sounds s-ö-r which acts as its signal", and that this "is the organizing principle for the whole of linguistics". The fact that the series of sounds which represents an idea such as 'sister', and that one cannot derive the meaning of that series of sounds without learning the language, is strong evidence for this principle. Nonetheless, there is evidence for a variety of situations where sound does create, affect or modulate meaning; words which exhibit such a non-arbitrary relationship are often called iconic. One of the most famous experiments which shows evidence for iconicity in language is that of the Bouba-kiki (BK) effect, in which participants associate round and spiky shapes with certain phonetic segments, as an iconic 'ground' creates a cross-modal sound-to-shape mapping (Ramachandran & Hubbard, 2001; Ahlner & Zlatev, 2010). The BK effect makes use of nonword labels, but iconicity is also found in natural language, such as in ideophones, a marked class of words with sensory meanings that are found in many non-European languages (Dingemanse, 2012). An English example of an ideophone would be *zigzag*, where the sound mimics the alternating movement it signifies. With respect to iconicity in the English language, Sidhu et al. (2021) showed that the phonemes associated with roundness/spikiness in the BK effect were more common in English words referring to round/spiky objects, while Winter et al. (2017) used native speaker ratings to show that words with sensory meanings are more iconic (for example, *hissing* was judged to be highly iconic, as onomatopoeia often are, whereas abstract terms like *permission* were judged not iconic).

Research on iconicity has shown that certain phonemes have iconic qualities in certain semantic contexts (see discussion below). As such, this study intends to further the research on iconicity by making use of natural language processing tools and native speaker iconicity ratings (from Winter et al., 2023) to discern the extent to which information about iconicity is contained in phonetic and semantic word embeddings. To this end, we employ linear regression models using phonetic and semantic word embeddings as the explanatory variables and iconicity ratings as the response variable. We expect there to be a degree of iconicity information in both the semantic and phonetic embeddings, given the aforementioned relationship between certain phonemes, semantic contexts and iconicity. We find that these models do indeed have a certain degree of success in predicting iconicity ratings, and that the trends in their predictions are consistent with other research on iconicity.

Background

Form and Meaning in Iconicity

The existence of iconicity in language is predicated on the idea that the form of a linguistic sign can be linked in some way to its meaning (that is, that a word can sound like what it means), and as such the question of how form is linked to meaning is fundamental to the study of iconicity. The phonetic form of a word cannot wholly determine its meaning, the fact that foreign languages are uninterpretable to us is just one indicator of this, and thus iconicity must function in a different way. Ahlner & Zlatev (2010) make use of Peircian semiotics to explain the functioning of iconic signs. In this framework, a *representamen* (or sign) represents an *object* to an *interpretant* (that is, the person interpreting the sign). The idea which links the sign to the object is called the *ground* (see Figure 1). Peirce identified three 'ideal types' of signs, which are differentiated by their type of *ground* (Peirce, 1992). In an iconic sign, the nature of the ground is that of similarity: the sign and the object share some similar qualities. Indexical signs are based on relations in time and space, but are not relevant to the current discussion, while symbolic signs are based on a ground of convention, and thus are typical of the archetypal Saussurean 'arbitrary' sign. Ahlner & Zlatev (2010) point out that, in reality, signs often consist of facets of these three 'ideal types', rather than being a pure example of any of them. These three types of signs (arbitrary, indexical and iconic) correspond to what Ferrara & Hodge (2018) argue are the three methods of signalling that make up our use of language: describing, indicating and depicting. These three methods are often used in tandem, in what are termed 'multi-modal composite utterances', which are typical of both spoken and signed languages (Enfield, 2009).

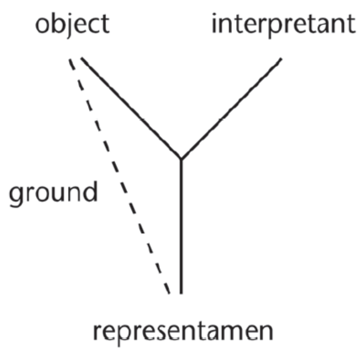

Figure 1: Illustration of Peircian semiotics (from Ahlner & Zlatev, 2010). The representamen (sign) is connected to the real word object by a ground. In the case of iconic signs, this ground is one of similarity.

The Bouba-kiki Effect and Cross-modal Iconicity

Iconic signs, therefore, are based on similarity, and this is perhaps most clear in the case of onomatopoeia, where the sound of a word such as *boom* resembles that of the explosion it represents. However, iconicity in language is not limited to sound-to-sound similarity; linguistic signs can also be based on similarity across modalities, such as a similarity between sound and shape. Perhaps the most famous example of cross-modal iconicity is a phenomenon known as the *bouba-kiki effect* (BK effect)*,* first reported as the *maluma-takete effect* in Köhler's *Gestalt psychology* (1929), where participants shown round and spiky shapes, such as those in Figure 2, consistently associated the name *maluma* with a round shape and *takete* with a spiky shape. The BK effect was popularised by Ramachandran & Hubbard (2001) and many variations of the original experiment have been carried out. Research on the BK effect shows that the majority of people perceive a relationship between curvilinear shapes and sonorants (such as /l/, /m/, /n/) and also voiced stops (most commonly /b/), while rectilinear shapes are associated with voiceless stops (like /p/, /t/, /k/) (Sidhu et al., 2021). This phenomenon has been found to hold across cultures and writing systems (Cwiek et al., 2022), with the notable exception of individuals with autism spectrum disorders

(Occelli et al., 2013), which some have linked to a lack of cross-modal integration (Gold & Segal, 2017). There is also evidence that this cross-modal mapping occurs prior to conscious awareness of the visual stimuli (Hung et al., 2017). In the case of the BK effect, interpretants perceive a ground of similarity between these shapes and sounds, which Ramachandran & Hubbard argued, in the case of *kiki,* is because “the sharp changes in visual direction of the lines in the [...] figure mimics the sharp phonemic inflections of the sound kiki, as well as the sharp inflection of the tongue on the palate” (2001: 19). However, there is also evidence that this ground is based in the acoustic properties of physical objects, as “round items are mathematically bound to produce, when hitting or rolling on a surface, lower-frequency spectra and more continuous sounds than same-size spiky objects” (Fort & Schwartz, 2022). Whatever the specific mechanism by which this ground is established, it is clear that the BK effect is an example of the universality of cross-modal iconicity.

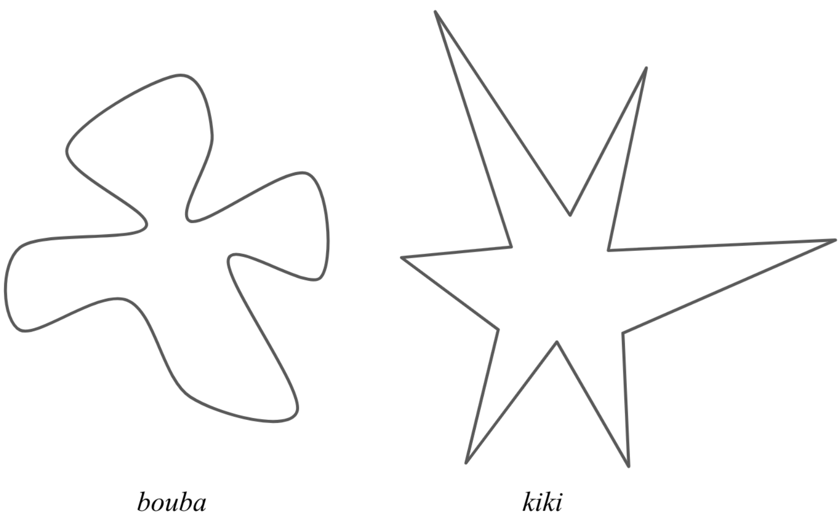

Figure 2: Examples of *bouba/kiki* type stimuli. The round shape on the left is most often associated with the nonword name *bouba* while the spiky shape is associated with *kiki*, suggesting an iconic relationship between the sound of the words and the two shapes.

Iconicity in Natural Language

While the BK effect is evidence of a cross-modal iconic relationship between phonemes in nonwords and certain shapes, such iconic relationships can also be found in natural languages to varying degrees. Macrolevel studies have shown that genealogically and geographically diverse languages persistently show sound-meaning association biases (Blasi et al., 2016), and neural networks have been used to show that a "certain amount of cross-linguistic correspondence between form and meaning is stable across languages", which can be used by the models to predict a word's meaning in an unseen language (de Varda & Strappavara, 2022: 14). Iconicity in the lexicon goes beyond onomatopoeia: ideophones, defined as "marked words which depict sensory imagery" (Dingemanse, 2012: 655), make use of iconic mapping of form and are found in many languages, although they are most prominently found in certain non-European languages, such as Japanese, the Bantu languages of sub-saharan Africa, and Basque, among many others (Dingemanse, 2018). There is a significant body of experimental work on Japanese ideophones in particular, which shows that participants with no knowledge of Japanese are able to guess the meaning of these words with an accuracy significantly above chance levels (Lockwood & Dingemanse, 2015), an indicator of the communicative value of iconicity in language. Dingemanse et al. (2015: 603) argue for a conception of natural language vocabulary structure "in which arbitrariness is complemented by iconicity (aspects of form resemble aspects of meaning) and systematicity (statistical regularities in forms predict function)", and that these differing form-meaning relationships are utilised with different functions.

Iconicity is also, of course, present in the English language, where ideophones such as *zigzag*, *roly-poly* and *willy-nilly* can be given as examples (Tedlock, 1999). Sidhu et al. (2021) carried out an experiment in which participants were asked to quantify whether 1,757

English object nouns were round or spiky, and found that phonemes associated with roundness and spikiness in the BK effect were more common in round and spiky English nouns, showing the existence of this same iconic phenomenon in the English language. Winter et al. (2023) collected iconicity ratings for 14,000 words and found that iconicity ratings were strongly correlated with sensory experience ratings, suggesting that English sensory words are more likely to be iconic. Phonaesthemes, sound-meaning correspondences which occur in different words, are present in English as in other languages. Examples include the association *gl-* with bright light *(glitter, gleam, glow, glare, glisten)*, or *sl-* with a falling or sliding movement *(slide, slump, slouch, slip).* Phonaesthemes are an example of systematicity, but are not necessarily iconic, as this depends on whether there is a similarity between form and meaning (Kwon & Round, 2015).

The Functional Roles of Iconicity in Language

Studies have shown that iconicity makes a significant contribution to word acquisition, language processing and language evolution (Winter et al., 2023). Imai & Kita (2014: 10) proposed the *sound symbolism bootstrapping hypothesis* which claims that iconicity facilitates word learning in children, as “pre-verbal infants detect sound symbolism in unfamiliar words and process them as if they were real words, which may lead them to (or solidify) the realization that speech sounds have meanings”. Further evidence for this is given by the overrepresentation of iconic words in child vocabularies, with children learning iconic words earlier and adults tending to use more iconic words when speaking with children (Perry et al., 2018). The role of iconicity in language evolution is exemplified by research that shows that, in the absence of a common language, individuals make use of gestures and iconic vocalisations in order to communicate successfully (Silva, 2020).

An important element in the discussion of iconicity is the subjective nature of the phenomenon. Iconicity is subjective because individuals differ in their interpretation of how iconic forms map to their referents (Barnes, 2023). This subjectivity is consistent with the centrality of the interpretant in the framework of Peircian semiotics, as previously discussed. The subjective nature of iconicity is reflected by the fact that certain sounds have been shown to be associated with different meanings depending on the semantic context; for example, the phoneme /i/ has been associated with small size, bitterness, angular shapes and brightness (Winter et al., 2023). This study makes use of native speaker iconicity ratings, which tap into the subjective nature of the phenomenon. Although such ratings are not able to provide information on the form-meaning mappings that occur in any given word, they do allow us to gain insight into the distribution of iconicity in a vocabulary as a whole (Winter & Perlman, 2021). The relationship between iconicity and concreteness has also been explored, with Lupyan & Winter (2018: 6) arguing that "iconicity limits abstraction and abstraction limits iconicity", given that iconic words tend to be less abstract, and that abstract words that are iconic tend to be perceived as more concrete.

In this study, semantic word embeddings will be employed to provide semantic information about words, while phonetic embeddings will be used to provide phonetic information. Semantic embeddings contain distributional information about the words they represent and the use of these embeddings is grounded in the Distributional Hypothesis, which asserts that similarity in meaning corresponds to similarity in linguistic distribution (Boleda, 2020). Phonetic embeddings, meanwhile, contain information about the articulatory features of a given word, such as whether a vowel is sonorant or a consonant voiced (Mortensen et al., 2016). The primary objective of this paper therefore is to ascertain the degree to which information pertaining to iconicity is encapsulated within both phonetic and semantic word embeddings through the use of linear regression models. By leveraging these resources, the study intends to deepen our understanding of the relationship between iconicity, phonetics, and semantics.

Methods

Iconicity Ratings

This study makes use of the iconicity rating dataset from Winter et al. (2023). The dataset is the result of an experiment where 1400 American English-speaking participants were given a definition of iconicity along with examples of iconic words, including *screech,* where the word sounds like the sound it represents, and *twirl* and *ooze*, which are examples of cross-modal iconicity, where there is a similarity between the the way the words sound and the movements they represent. The participants rated English words on a Likert scale from 1, not iconic at all, to 7, signifying a very iconic word (see Figures 3 and 4 for an example of a trial and the rating distribution). A previous edition of this dataset (Winter et al., 2017) made use of negative iconicity ratings where participants identified words which sounded like the opposite of what they meant. However, the 'anti-iconic' end of the scale was disposed with in this newer iteration, as it was found that participants did not use the negative scores nearly as much and were more inconsistent when they did, likely due to conceptual confusion (Winter et al., 2023).

The dataset used in this study contains a total of 14,776 words along with their mean iconicity ratings. For the purposes of modelling, the iconicity ratings were rescaled to 0 to 1. See Table 1 for an overview of the words with the highest and lowest iconicity ratings.

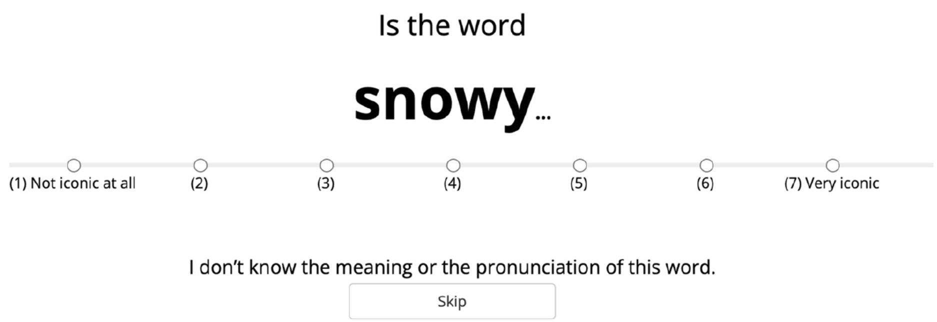

Figure 3: An example trial as presented to raters (from Winter et al. 2023).

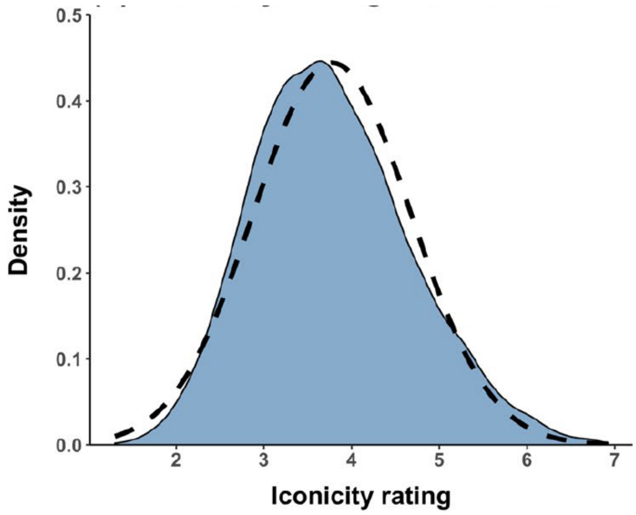

Figure 4: Kernel density plot of average iconicity rating distribution; the dashed line indicates a normal distribution with the same mean and standard deviation (from Winter et al., 2023).

Word	**Iconicity rating**
oomph	0.987
swish	0.985
wiggle	0.983
creak	0.967
clunk	0.967
…	
gnome	0.067
are	0.061
if	0.05
partial	0.05
how	0.05

Table 1: Illustration of iconicity dataset from Winter et al. (2023), rescaled from 0 to 1.

Phonetic Embeddings

As a measure of the phonetic content of the words in the dataset, the Epitran-Panphon pipeline was used to obtain phonetic vectors for each token. Epitran (Mortensen et al., 2018) is used to convert tokens into phonemic representations using the International Phonetic

Alphabet (IPA), while Panphon (Mortensen et al., 2016) converts the IPA representations into articulatory feature vectors per segment. The articulatory features are represented by vectors with 24 dimensions with values of -1, 0 and 1, with -1 and 1 referring to - and + for each feature, as in [±consonantal], while 0s occur in cases where that particular feature cannot apply. Since both -1 and 0 refer to an absence of the given feature, they were both treated as 0 for the purposes of modelling. The tokens in the dataset vary in their number of segments: for example, *shh* (/ʃ/) has just one segment while *retina* (/ɹɛtənə/) has six. Since a linear regression model requires the same number of variables for each datapoint, the mean of the segmental features was taken as a general representation of the articulatory information for each token. However, this does mean that sequential information is lost. Another approach was trialled, where the length of each vector sequence was normalised through the addition of empty vectors, but this failed to produce useful results. In order to narrow down the features to be used for linear regression, a correlation matrix was used as an indicator of how the mean phonetic information is distributed among the words in the dataset. As Figure 5 shows, a number of features had no values, specifically the spread/constricted glottis, velaric, long, high tone and high register features (10, 11, 20, 22, 23 and 24 respectively); this is very likely because these features are not present or unspecified in English. These features were then eliminated for the purposes of the linear regression, since they are not informative. Figure 5 also shows there are not any strong correlations between the remaining variables, and since our primary goal is to see whether such feature-based representations are predictive of iconicity, we included all informative dimensions. For more information about the phonetic information captured by the Panphon features, see the Appendix.

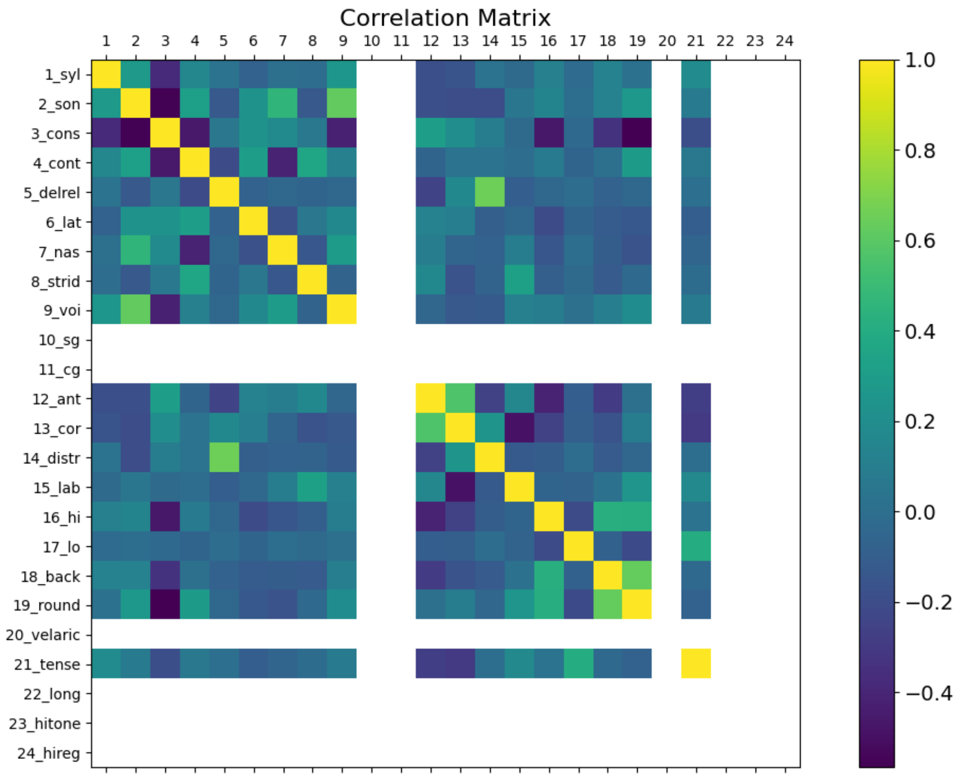

Figure 5: Panphon Pearson's correlation matrix for mean phonetic representations of words in the dataset.

Semantic Embeddings

As a representation of the semantic content of each word, Fasttext pre-trained word embeddings for English were used (Mikolov et al. 2017). These embeddings consist of 300 dimensions, and the number of dimensions needed to be reduced in order to obtain optimal results. Linear regression models with very high numbers of features compared to the number of samples in the dataset can run into issues such as overfitting and a lack of interpretability, although in this case the semantic features are not directly interpretable in the first place, along with increased computational costs. Principal Component Analysis (PCA), as implemented in the Scikit-learn Python package (Pedregosa et al., 2011), was used in order to reduce the number of dimensions. In order to ascertain the most informative number of dimensions for the linear model, preliminary linear models were run with 1-300 dimensions, each using a randomised 70/30 train/test data split to avoid overfitting, and their root mean square errors (RMSE) were taken as a measure of their performance. The results of this step are shown in Figure 6. 185 dimensions were used for the final model as this model had the lowest RMSE. Additionally, in order to assess the relationship between the semantic model's predictions and concreteness/abstractness, the concreteness ratings from Brysbaert et al. (2014) were added for each word. These ratings were also based on native speaker intuitions, where participants were asked to rate words on a scale of 1 (abstract) to 5 (concrete), and these ratings were also rescaled to 0 to 1 for the purposes of analysis.

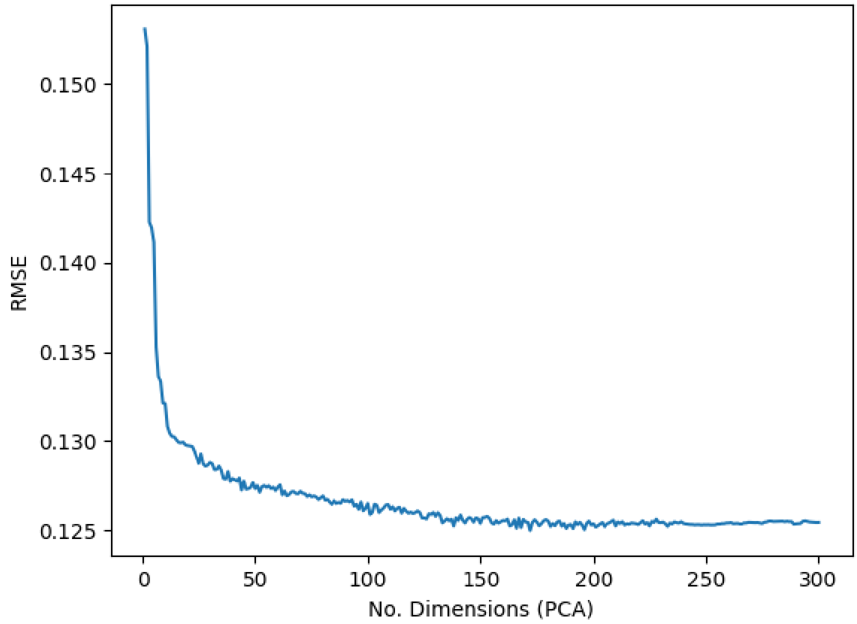

Figure 6: Semantic vector models - number of dimensions (x-axis) vs. RMSE (y-axis).

Taking stock, the iconicity dataset, phonetic vectors and semantic vectors were used to create three linear models: a **phonetic model** using the 18 phonetic features as independent variables, a **semantic model** using the 35 dimensions from PCA as independent variables, and a **combined model** which concatenated both the phonetic and semantic features for the independent variables. In all three cases the iconicity score was the dependent variable. The linear regression models were fit with the Scikit-learn Python package (Pedregosa et al., 2011). R^2, mean absolute error (MAE), mean square error (MSE) and root mean square error

(RMSE) were taken as performance metrics for each model. The Kendall's **τ** score was taken as the rank correlation between the word list ordered by each model prediction measured against the same list ordered by iconicity rating, and also as the rank correlation between the semantic model's predictions and the aforementioned concreteness ratings.

The Python code for preprocessing and analysis can be found at: https://github.com/jotadwright/iconicity_embeddings

Results

The performance metrics for each model can be found in Table 2. It should be noted that higher numbers of variables artificially inflate the R^2 score, which may be the case for the semantic and combined models (which have 185 and 203 variables respectively). The Kendall's **τ** score is given for the rank correlation with iconicity ratings for all model predictions, while the Kendall's **τ** score for correlation with concreteness is just given for the semantic model.

Model	**R^2**	**MAE**	**MSE**	**RMSE**	**Kendall's τ (correlation with iconicity)**	**Kendall's τ (correlation with concreteness)**
Phonetic	10.38	0.113	0.02	0.141	0.211	-
Semantic	34.11	0.097	0.015	0.121	0.122	0.103
Combined	11.55	0.113	0.019	0.141	0.222	-

Table 2: Performance metrics for the three models

Table 3 contains the coefficients for each feature in the phonetic model. Since the semantic dimensions are not directly interpretable, this has been omitted for the semantic and combined models.

Feature	**Coefficient**
[±syllabic]	-0.187
[±sonorant]	0.152
[±consonantal]	0.433
[±continuant]	0.057
[±delayed release]	0.045
[±lateral]	-0.083
[±nasal]	-0.204
[±strident]	-0.13
[±voice]	-0.002
[±anterior]	0.054
[±coronal]	-0.044
[±distributed]	0.146
[±labial]	0.077
[±high]	0.162
[±low]	0.059
[±back]	0.027
[±round]	0.048
[±tense]	0.156

Table 3: Phonetic model feature coefficients.

Tables 4, 5 and 6 show which words each model predicted as most and least iconic, which serve to illustrate the trends in words which each model selects as iconic or not iconic. These tables contain the iconicity ratings (from Winter et al., 2023) and the model predictions for comparison, and Table 5 also contains the concreteness ratings from Brysbaert et al. (2014). Both of these rating sets have been rescaled to 0 to 1.

Word	**IPA**	**Iconicity**	**Phonetic model prediction**
shh	ʃ	0.9	0.667
hooch	hut͡ʃ	0.621	0.645
booth	buθ	0.35	0.641
who	hu	0.433	0.641
hoop	hup	0.611	0.64
cool	kul	0.717	0.634
clue	klu	0.208	0.634
glue	glu	0.6	0.634
ghoul	gul	0.667	0.634
puke	puk	0.742	0.634
heir	ɛɹ	0.283	0.312
err	ɛɹ	0.85	0.312
air	ɛɹ	0.383	0.312
retina	ɹɛtənə	0.394	0.308
Aurora	ɹ̟ɔɹə	0.396	0.29
area	ɛɹiə	0.233	0.289
enema	ɛnəmə	0.136	0.277
aura	ɔɹə	0.394	0.273
era	ɛɹə	0.3	0.248
eh	ɛ	0.773	0.094

Table 4: Phonetic model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

Word	IPA	Iconicity	Semantic model prediction	Concreteness
pooch	puť͡ʃ	0.517	0.837	0.83
mouse	maws	0.597	0.832	0.958
huge	hjud͡ʒ	0.583	0.818	0.635
clipped	klɪpt	0.6	0.799	0.602
igloo	ɪglu	0.533	0.784	0.933
coop	kup	0.5	0.783	0.865
sooth	suθ	0.727	0.783	0.277
cheesecake	t͡ʃizkejk	0.667	0.78	0.992
camp	kæmp	0.45	0.772	0.837
blues	bluz	0.683	0.771	0.328
abandoned	əbændənd	0.533	0.233	0.38
evangelicalism	ɛvænd͡ʒɛlɪkəlɪzəm	0.385	0.228	0.125
fantasize	fæntəsajz	0.5	0.227	0.285
sufficiency	səfɪʃənsi	0.417	0.226	0.202
unfulfilled	ʌnfʊlfɪld	0.617	0.225	0.157
repair	ɹɪpɛɹ	0.45	0.225	0.58
fern	fɹ̩n	0.367	0.224	1
dwell	dwɛl	0.744	0.221	0.405
erotica	ɹ̩ɑtɪkə	0.583	0.221	0.57
pain	pejn	0.576	0.196	0.625

Table 5: Semantic model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

Word	IPA	Iconicity	Combined model prediction
shh	ʃ	0.9	0.694
booth	buθ	0.35	0.662
hoop	hup	0.611	0.655
hooch	hut͡ʃ	0.621	0.653
puke	puk	0.742	0.652
who	hu	0.433	0.648
coup	ku	0.4	0.647
sooth	suθ	0.727	0.645
gloom	glum	0.867	0.643
clank	klæŋk	0.894	0.643
nausea	nɔziə	0.533	0.312
etcetera	ɛtsɛtɹ̩ə	0.633	0.312
retina	ɹɛtənə	0.394	0.304
Aurora	ɹ̩ɔɹə	0.396	0.301
air	ɛɹ	0.383	0.296
enema	ɛnəmə	0.136	0.293
aura	ɔɹə	0.394	0.29
area	ɛɹiə	0.233	0.272
era	ɛɹə	0.3	0.241
eh	ɛ	0.773	0.119

Table 6: Combined model predictions. Darker hues indicate values closer to 1 while lighter hues indicate those closer to 0.

Discussion

The performance metrics shown in Table 2 indicate that the semantic model performs best overall, with lower error measures and a higher R^2. The phonetic model does worse on these metrics, but it does have a better Kendall's τ score, which suggests that the model's ranking of words from most to least iconic is more similar to that of the human iconicity ratings. The combined model performs very similarly to the phonetic model, making a very slight improvement in the metrics, suggesting that the combined model prioritises the phonetic variables over the semantic ones.

The phonetic model coefficients suggest that high proportions of consonantal, sonorant and distributed segments are positively correlated with iconicity, while nasal, strident and syllabic segments are negatively correlated. However, examination of the top and bottom predictions for the phonetic model perhaps give a more concrete picture of the kind of phonemic structures the phonetic model prefers: we can see the repetition of the phonemes /u/, /k/ and /h/, and a trend of generally short words. Examining the words that the phonetic model deems least iconic we can see the repetition of the phonemes /ɛ/, /ə/ and /ɹ/. The top predicted word, /ʃ/, is a pure consonant, in line with the high weight the model gives the consonantal variable. The fact that /u/, a rounded back vowel, occurs so frequently in the top predictions is interesting given that the rounded and back features were not given particularly high importance by the model. This perhaps suggests that interaction between the different phonetic variables is rather complex, and the averaging of the scores over a word likely also contributes to this. Nonetheless, the model does seem to consistently select certain segments, which results in some relatively accurate approximations of high-iconicity words (*shh*, *cool*, *hooch*) with some phonetically similar errors (*clue, booth*). There seems to be some convergence with Sidhu et al.'s (2021: 1395) examination of roundness and spikiness in

English nouns, where the phoneme /u/ was found to be iconically related to roundness, while /ʃ/, /t͡ʃ/ and /k/ were all linked to object spikiness. All of these phonemes occur among the top ten words as predicted by the phonetic model, /u/ and /k/ repeatedly. /k/ and /u/ were also found in a number of strong worldwide sound-meaning associations in Blasi et al. (2016). All in all, the metrics for the phonetic model show that variation in the averaged phonetic vectors can explain variation in iconicity rating to some extent, although this relationship is perhaps not as strong as for the semantic model, and analysis of the predictions shows some concurrence, in terms of phonemes proven to have iconic properties, with other research.

The combined model performs very similarly to the phonetic model as the performance metrics show. The combined model's top predictions also reveal great similarity with the phonetic model, as we can see a repetition of the phonemes /u/, /k/ and /h/ among the top predictions and of the phonemes /ɛ/, /ə/ and /ɹ/ among the bottom predictions. The two models are not identical, there are differences in the predicted iconicity they give for many words, but the overall picture indicates that the combined model is very strongly influenced by the same variables which are included in the phonetic model.

The semantic model, however, clearly gives very different results. The coefficients for the semantic model are derived from pre-trained word embeddings, reduced to a smaller number of dimensions via PCA, and it is not possible to directly interpret them, as the dimensions themselves do not refer to semantic properties but instead are used conjunctively to locate words in a semantic space (Boleda, 2020). The top predictions for the semantic model do not generally seem to be semantically related. Nevertheless, comparing the top and bottom predictions does suggest that this model selects concrete words over abstract words, with the top words containing concrete terms such as *mouse, clipped* and *huge*, while abstract concepts such as *abandoned, fantasize* and *evangelicalism* are found with low iconicity predictions. A notable exception to this trend is *fern*, with a low iconicity prediction and high concreteness rating. This trend is reflected in the Kendall's rank correlation between the semantic model's iconicity predictions and concreteness ratings for the same words. The score reflects a positive, although relatively weak, correlation between the semantic model's predictions and concreteness ratings, meaning that the model generally selects words with higher concreteness ratings as being more iconic. Given that the semantic model learns from the distributional information in semantic word embeddings, it makes sense that the model would pick up on the difference in use of abstract and concrete words. This is in line with Lupyan & Winter's argument that "iconicity limits abstraction and abstraction limits iconicity" (2018: 6), as it is generally harder to perceive similarity between a sound and an abstract concept than, say, a shape. However, it should be noted that there are some contradictory findings with respect to concreteness and iconicity, with concreteness ratings found to negatively correlate with iconicity ratings in some studies (Winter et al., 2023). Surprisingly, the most obvious element that almost all the words in the top semantic predictions share is the phoneme /u/, suggesting that the model has somehow indirectly selected words with this phoneme. Word length also seems to be a factor, with the top ten semantic model predictions having a mean length in number of segments of 3.8, while the bottom ten has an average of 8. Overall, the semantic model has the strongest performance metrics, although the rank correlation shows that it is not as strong at ordering the words from highest to lowest iconicity, and analysis of the model's predictions shows that it is likely tapping into the aforementioned concreteness/abstractness phenomenon.

The purpose of this study was to identify whether there is information about iconicity in phonetic and semantic word embeddings through the use of linear regression models and native speaker iconicity ratings. The results of these models show that these embeddings do have some explanatory value with respect to iconicity ratings. However, these findings have some limitations. Firstly, iconicity ratings such as those used in this study (from Winter et al., 2023) are based on native speaker intuitions. This is in part a strength, as iconicity is subjective by nature, yet there is a degree of consistency in this phenomenon, in the way in which users of a language identify iconicity across the lexicon, and taking average ratings can capture this (Winter & Perlman, 2021). However, these ratings do no at all explain why any given word is iconic or not, and do not give information about specific form-meaning relationships. What's more, the issue of the extent to which participants confuse the concept of iconicity with other ideas, such as semantic transparency, has been raised with respect to the results of such iconicity rating experiments. Dingemanse & Thompson (2020) pointed out that compound words such as 'dishwasher' and 'skateboard' received high iconicity ratings in Perry et al. (2017), yet are not actually iconic, which they attributed to the phrase *a word sounding like what it means* causing a confusion of semantic transparency and iconicity. As such, there is a possibility of a bias towards such words having inflated iconicity ratings in datasets like the one used for this study. Similar criticisms have been made over the reliability of the concreteness ratings used in the analysis of the semantic model's predictions, such as that the mean ratings in the middle sometimes reflect where there is disagreement between participants (Pollock, 2018). Another limitation is in the nature of semantic word embeddings, which locate words in vector spaces where the semantic distance between words can be measured. However, the dimensions that make up semantic word embeddings do not refer to any specific semantic properties, and therefore it is not possible to extrapolate, for example, what semantic content caused the semantic model to rate *pooch* as more iconic than *independence,* from the word embeddings themselves. Nonetheless, the embeddings and ratings do have explanatory value for the purposes of this research, and the convergence of the results presented here with preexisting iconicity research supports this.

Further research on this subject could build upon these findings with multi-modal word embeddings, which are enhanced with perceptual input and could be exploited to explore the cross-modal nature of iconicity in language (Verő & Copestake, 2021). Words which have sensory meanings are often highly rated for iconicity (Winter et al., 2017), and incorporating perceptual input may help improve the accuracy and robustness of regression models used to predict iconicity ratings. This study also focuses on iconicity ratings in English, and similar work in other languages would be valuable and create opportunities for cross-linguistic analysis. Panphon has functionality for a number of different languages, semantic word embeddings for a large variety of languages are also widely available, and iconicity ratings are being developed for other languages too (for example, Hinojosa et al. (2021) for Spanish).

Conclusion

This study explored the phenomenon of iconicity in language and its relationship to phonetics and semantics. The concept of iconicity challenges the conventional Saussurean notion of the arbitrariness of linguistic signs, suggesting that there can be a connection between the form of a word and its meaning. We examined the different types of signs in Peircian semiotics, highlighting the role of similarity as the ground for iconic signs and the cross-modal nature of many iconic signs. The Bouba-Kiki effect, which demonstrates a cross-modal sound-to-shape mapping, and the presence of ideophones in various languages are examples of iconicity in action. The subjective character of iconicity was emphasised, as individuals may interpret the mappings between form and meaning differently, and the study utilised native speaker iconicity ratings which capture this subjective aspect.

We aimed to contribute to the existing research on iconicity by utilising natural language processing tools and native speaker iconicity ratings. By employing linear regression models with phonetic and semantic word embeddings as explanatory variables and iconicity ratings as the response variable, we sought to determine the extent to which information about iconicity is encoded in these embeddings. The findings of this study indicate that these models were able to predict iconicity ratings to a certain degree of success. Moreover, the trends observed in the predictions align with previous research on iconicity, particularly with certain phonemes repeatedly found in iconic words, and in the relationship between iconicity and concreteness. This research demonstrates the potential of incorporating natural language processing techniques and word embeddings in exploring the role of iconicity in language. Future studies could build upon these findings by utilising multi-modal word embeddings to further enrich the words' semantic representations, or by using the approach taken in this paper with other languages.

Bibliography

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38(1/4), 298-348.

Barnes, K. (2023). Subjectivity, perception and convention in ideophones and iconicity. *SKASE Journal of Theoretical Linguistics*, *20*(1).

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

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213-234.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46, 904-911.

Dingemanse, M. (2012). Advances in the cross‐linguistic study of ideophones. Language and Linguistics compass, 6(10), 654-672.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603-615.

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics*, 3(1).

Dingemanse, M., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. Language and Cognition, 12(1), 203-224.

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9, 716.

Fort, M., & Schwartz, J. L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports*, 12(1), 19172.

Gold, R., & Segal, O. (2017). The bouba-kiki effect and its relation to the Autism Quotient (AQ) in autistic adolescents. *Research in Developmental Disabilities*, 71, 11-17.

Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A., & Ferré, P. (2021). Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*, 53, 1262-1275.

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.

Kwon, N., & Round, E. R. (2015). Phonaesthemes in morphological theory. *Morphology*, 25, 1-27.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170137.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in psychology*, 6, 1246.

Magnus, M. (2013). A history of sound symbolism. *The Oxford handbook of the history of linguistics* 191–208. Oxford, England: Oxford University Press

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.0940*5.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 3475-3484).

Mortensen, D. R., Dalmia, S., & Littell, P. (2018). Epitran: Precision G2P for many languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, 64(1), 52-72.

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The Takete—Maluma phenomenon in autism spectrum disorders. *Perception*, 42(2), 233-241.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.

Peirce, C. S. (1992). On a new list of categories (1867). *The writings of Charles S. Peirce: a chronological edition*, 2, 49-58.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, 21(3), e12572.

Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, 50(3), 1198-1216.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. Journal of consciousness studies, 8(12), 3-34.

Saussure, F. M. (1972). *Course in general linguistics*. Open Court.

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, 28, 1390-1398.

Silva, V. M., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society open science*, 7(1), 182056.

Tedlock, D. (1999). Ideophone. *Journal of Linguistic Anthropology*, 9(1/2), 118-120.

de Varda, A. G., & Strapparava, C. (2022). A Cross‐Modal and Cross‐lingual Study of Iconicity in Language: Insights From Deep Learning. *Cognitive Science*, 46(6), e13147.

Verő, A. L., & Copestake, A. (2021). Efficient Multi-Modal Embeddings from Structured Data. *arXiv preprint arXiv:2110.02577*.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, 18(3), 443-464.

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard*, 7(1).

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled/r/is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, 12(1), 1035.

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2023). Iconicity ratings for 14,000+ English words. *Behavior Research Methods*, 1-16.

Appendix

[±syllabic]	Is the segment the nucleus of a syllable?
[±sonorant]	Is the segment produced with a relatively unobstructed vocal tract?
[±consonantal]	Is the segment consonantal (not a vowel or glide, or laryngeal consonant)?
[±continuant]	Is the segment produced with continuous oral airflow?
[±delayed release]	Is the segment an affricate?
[±lateral]	Is the segment produced with a lateral constriction?
[±nasal]	Is the segment produced with nasal airflow?
[±strident]	Is the segment produced with noisy friction?
[±voice]	Are the vocal folds vibrating during the production of the segment?
[±anterior]	Is a constriction made in the front of the vocal tract?
[±coronal]	Is the tip or blade of the tongue used to make a constriction?
[±distributed]	Is a coronal constriction distributed laterally?
[±labial]	Does the segment involve constrictions with or of the lips?
[±high]	Is the segment produced with the tongue body raised?
[±low]	Is the segment produced with the tongue body lowered?
[±back]	Is the segment produced with the tongue body in a posterior position?
[±round]	Is the segment produced with the lips rounded?
[±tense]	Is the segment produced with an advanced tongue root?

Panphon features used, descriptions taken from Mortensen et al. (2016)