Bidimensional Features of Prosody and Semantic are Sufficient for Inferring and Predicting Threatening Speech Comprehension

Simon Busch-Moreno*, David Vinson

Division of Psychology and Language Sciences, University College London

*Corresponding author: simon.busch.15@ucl.ac.uk

Abstract

The present study investigates how variation in acoustic measures and lexical semantic properties affect the comprehension of threatening speech. Two experiments were conducted in British English, where participants had to rate the threat of three types of sentences: sentences neutral in prosody and semantics, sentences containing only prosodic threat, and sentences containing only semantic threat. They also rated the sentences for arousal and valence, and categorised them by type of emotional expression. Threat ratings were analysed via Bayesian ordered-logistic regressions using acoustic measures and affective norms as regressors. Classification was assessed via simple Bayesian categorical (Dirichlet prior) models. Acoustic properties of stimuli were compared through a Bayesian method for comparison of means to test whether neutral stimuli differ, on average, from their threatening counterparts. Acoustic but not semantic properties affected ratings of threatening prosody (increased pitch and decreased voice quality predicting increased threat), while the reverse was true for ratings of threatening semantics (increased arousal and decreased valence predicting increased threat). Using data from the first speaker (first experiment) the models were able to predict ratings given to sentences produced by the second speaker (second experiment). Furthermore, threatening sentences of both types tended to be categorized as angry or enraged. These results are interpreted as supporting the sufficiency of the selected measures (pitch, voice quality, arousal, valence) to characterise threat. Although multidimensional models may be better suited for describing threatening prosody/semantics, the present approach demonstrates that selected bidimensional features are sufficient to identify and predict threat comprehension efficiently and accurately.

Keywords: prosody, semantics, speech, emotion, Bayesian models

Introduction

Language and speech processing models propose that slow-rate frequency (spectral) information is crucial for the processing of prosody, while fast-rate frequency is crucial for phonological-level processing (Poeppel et al., 2008; Zatorre et al., 2002). At the sentence level, this could imply that prosody (suprasegmental features) is influential for comprehending emotional meaning, even in the absence of emotional lexical items. A lexical item should drive the semantic emotional meaning of the sentence, but not the prosodic emotional meaning. In the present study we seek to find whether emotional sentences can be characterised by analogous prosody and semantic features. In particular, we want to focus on stimuli able to convey threat. By threat, we mean indirect/potential threat, such as sentences expressing anger or rage in the form of offensive or hurting language. In short, we want to know whether spectral variations of threatening sentences will be able to describe prosodic threat in a similar way as features (i.e. norms) of lexical items will be able to describe semantic threat. More precisely, we want to propose a bi-dimensional approach, where acoustic measures detecting increases in pitch and decreases in voice-quality should be sufficient to characterise prosodic threat. In an analogous manner, lexical norms describing increases in arousal and decreases in valence for sentences' content words should be sufficient to characterise semantic threat. The present aim is not to provide a complete model for describing threatening semantics or prosy, which may be composed of many interrelated features. Rather, the goal is to find a minimalistic approach, where the adjustment of two features in inverse proportion, such as increases in pitch vs decreases in voice quality, would provide a simple way of characterising threat production and predicting threat identification. Hence, the present model should be understood as derived from more complex previous models (e.g. Banse and Scherer, 1996; Hammerschmidt and Jürgens, 2007; Lausen and Hammerschmidt, 2020), in such a way that it retains inferential capacity (though not fully descriptive) and facilitates prediction. This could provide useful insights into emotional expression/comprehension, which may be extended from threat or 'hot anger' to other emotions.

Prosody is understood as the variation and combination of suprasegmental features, such as sound intensity, rhythm or pitch; combination and variation which, together with semantic content, can be intended for the communication of emotion (Lausen and Hammerschmidt, 2020). Thus, the comprehension of prosody might depend

upon contrasts between and presence/absence of several features, which have been adjusted by the speaker to convey a specific emotion (e.g. anger). This ‘tuning up’ of sorts can imply the increase of certain features and the decrease of others which together will imbue the stimulus (i.e. sentences) with emotional meaning. For instance, it has been observed that emotional meanings as different as joy and rage are similar in having higher pitch, and mainly differ in being less intense and having less energy at high frequency ranges (Hammerschmidt and Jürgens, 2007). For the case of threatening speech, this would imply that voice-quality measures focusing on frequency range differences, such as Hammarberg index which measures differences between 0-2000hz and 2000-5000hz ranges (Banse and Scherer, 1996), should be lower for raging or threatening sentences; while average or median fundamental frequency (pitch) should be higher. Whether increases in median pitch and decreases in Hammarberg index (voice-quality) are sufficient to explain increases in threat level (e.g. sentences rated in a Likert scale) is a relevant empirical question.

Semantic meaning can be understood, linguistically, as the meaning derived from lexical items or explicit sentence context (Kemmerer, 2015). Hence, if a stimulus has no meaningful variations in prosody, it could still convey emotionality through its content. Different from the slower rate at which prosodic meaning operates, phonemic articulations function as fast rate frequency changes which can be combined to produce lexical items and vary their meaning (Hickok and Poeppel, 2007). Hence, if the focus is on threatening semantics, threat could be characterised solely by lexical content. Thus, human ratings of words’ meanings, such as the emotion norms collected by Warriner et al. (2013), could provide a good means of quantifying the (semantic) threat level of lexical items, and may also quantify the threat level of sentences containing those lexical items: sentences containing words higher in arousal and lower in valence may be understood as angrier and more threatening. Consistent with this, it has been observed that words conveying threat (Ho et al., 2015) show high arousal and low valence values (see also Borelli et al., 2018 for similar patterns for pain-related words). Although sentences not containing emotional lexical items can convey emotional meaning (Lai et al., 2015), if emotionally loaded items are present, these will normally drive the emotional meaning of the sentence (e.g. Paulmann and Kotz, 2008; Yan and Sommer, 2018). This, let us propose another relevant empirical question, namely whether sentences containing lexical items that are high in arousal and low in valence will be understood as threatening.

Although arousal and valence at the lexical level are not necessarily the only features composing semantic threat, or pitch and voice-quality are not necessarily the only features composing prosodic threat, they may be sufficient to quantitatively characterise stimuli that are able to be felt and comprehended as threatening (offensive, harming, etc.). This should provide two comparable bidimensional maps of threatening semantics (arousal and valence dimensions) and prosody (pitch and voice-quality dimensions). For instance, some EEG studies have observed that valence and arousal variations in lexical items modulate a late positive complex (LPC) event-related potential (ERP) (Citron et al., 2013; Delaney-Busch et al., 2016; Imbir et al., 2017; Zhang et al., 2014). This reflects late-phase effects associated with higher order emotional/cognitive processing. Although these studies did not characterise arousal/valence variation in terms of specific emotions (e.g. anger), they do indicate that lexical meaning properties, such as valence and arousal norms, can modulate responses in a similar way to prosody-related acoustic properties. The latter has been observed by EEG experiments indicating that angry prosody (characterised mainly by increased pitch and intensity) can modulate ERP activity, including LPC (Burra et al., 2018; Chen et al., 2011; Pell et al., 2015). We propose that voice-quality is a required complement to pitch measures, as increases in pitch alone could be also related to other emotional expressions (Hammerschmidt and Jürgens, 2007). Moreover, we propose that semantic (arousal and valence) and prosodic (pitch, voice-quality/roughness) features can be understood as analogous. This may be important for comparing emotional (threatening) stimuli and understanding their effects on listeners. We emphasise that this is not a one-to-one relationship, namely is not that pitch is directly analogous to arousal and voice-quality to valence, but rather this is an analogy between the simultaneous increase/decrease of arousal/valence and pitch/voice-quality.

In the following experiments we test whether this analogous relationship holds, testing whether median pitch and Hammarberg index can characterise threatening prosody, vs. valence and arousal characterising threatening semantics. This leads to the following operative hypothesis: threatening prosody is higher in pitch and rougher (high median pitch, low Hammarberg index), and threatening semantics are arousing and negative (high arousal, low valence). This allows the following predictions to be made: 1) Sentences that contain only threatening prosody (neutral semantics and threatening prosody: termed “Prosody” henceforth) should not be characterised by arousal and

valence compared to sentences that are neutral in both semantics and prosody (“Neutral”), but clearly characterised by median pitch and Hammarberg index. 2) Sentences that contain only threatening semantics (neutral prosody and threatening semantics: “Semantic”) should not be characterised by pitch and roughness compared to entirely neutral sentences, but clearly characterised by arousal and valence. 3) Therefore, sentences containing lexical items which are high in arousal and low in valence should be easily categorised as semantically threatening, but prosodically neutral. 4) Similarly, sentences recorded as threatening (high in median pitch and low in Hammarberg index) should be easily categorised as prosodically threatening but semantically neutral.

To test these hypotheses and predictions, Prosody stimuli were selected using content words with arousal and valence around the median (or at least median arousal and higher valence), and recorded by asking two London English speakers (untrained in acting) to produce an aggressive voice (as if threatening someone). Semantic stimuli were selected to contain words high in arousal and low in valence and the speaker was requested to record them in a neutral voice (as if conversing with someone). Neutral stimuli were selected to be around the median in arousal and valence (or at least median arousal and higher valence) and requested to be spoken in a neutral voice (as if conversing with someone). For each of the two speakers, ten participants rated semantic threat of all three types of sentences (how threatening the sentence meaning is), and ten different participants rated prosodic threat (how threatening the speaker sounds). This allows to test the effects of sentences’ features on ratings, and also to conduct a cross-speaker comparison (i.e. use ratings of the first speaker to predict ratings of the second speaker). In order to test whether sentence-level measures of valence and arousal were adequately predicted by word level measures, participants also rated sentences’ arousal and valence using self-assessment manikin (SAM) scales as used by Warriner et al. (2013), again focusing either on the sentences’ meaning or sound. Finally, they were asked to judge each sentence’s emotion by adding them to one of four emotional categories (calm, neutral, angry, enraged).

Methods

Participants

For session 1 (Speaker 1's sentences), 10 UCL students (9 females, mean age = 19.1, all cisgender) rated auditorily presented neutral and prosody-threat sentences (Prosody experiment), and 10 University College London (UCL) students (9 females, mean age = 19.5, all cisgender) rated auditorily presented neutral and semantic-threat sentences (Semantic experiment); all students received coursework credits for their participation. For session 2 (Speaker 2's sentences), 10 voluntary participants (6 females, mean age = 20.9, all cisgender) rated auditorily presented neutral and prosody-threat sentences (Prosody experiment), and 10 voluntarily participants (4 females, mean age = 21.6, all cisgender) rated auditorily presented neutral and semantic-threat sentences (Semantic experiment). All participants were informed of the potentially offensive nature of stimuli and of GDPR regulations regarding data protection and gave their consent to participate. Participants declared not having auditory, psychiatric or neurological problems, and speaking English as first or native language.

Materials

Sentences are divided in four categories, Congruent (threatening prosody and threatening semantics), Neutral (neutral prosody and neutral semantics), Prosody (threatening prosody and neutral semantics), Semantic (neutral prosody and threatening semantics). Sentences for session 1 were selected from (transcribed) movie subtitles based on the following criteria. Sentences containing any word over 5 points in Warriner et al. (2013) arousal scale, and below 4.5 points in the valence scale was considered semantically threatening (both scales range from 1-9). Every word with less than 5 arousal points and between 4.5 and 6 valence points was considered semantically neutral. Only sentences with Zipf's log frequencies over 3 were included. Sentences for session 2 were manually created in a fixed phrasal structure of subject (second person pronoun), verb (to be), and adjective or noun (threatening word); an indefinite article preceded the noun if required (e.g. 'You're despicable', 'You are an animal'). Offensive adjectives were chosen to be words high in arousal (over 5 points) and low in valence (below 5 points), with just two exceptions (i.e., animal, pig) where this specific sentence context substantially altered the default interpretation of a word in isolation; neutral adjectives were chosen to be as low as possible in arousal (below 6) and as close

to median as possible in valence. Please note that arousal/valence thresholds for session 2 sentences were more lenient, because the simplified and controlled phrasal structure selected for this experiment (i.e. You are a [lexical item]) made it difficult to find enough appropriate lexical items. Sentences for session 1 were produced by Speaker 1, a male speaker from London, with English as his first language (upper-middle class London accent). Sentences for session 2 were produced by Speaker 2, a male speaker from London with English as his first language (lower-middle class London accent). Speakers recorded in separate recording sessions, and were instructed to speak to the microphone in a threatening or aggressive voice (as if threatening someone) for recording Congruent and Prosody sentences, and in a neutral voice (as if conversing with someone) for Neutral and Semantic sentences. Recordings were performed in an isolated anechoic chamber using a RODE NT1-A1 microphone. This resulted in 54 sentences per category for rating session 1 (Speaker 1) and 40 sentences per category for rating session 2 (speaker 2). These materials were recorded as stimuli for further experiments (Busch-Moreno et al., in press). We also recorded Congruent sentences (both semantic and prosodic threat) for the purpose of further experimental work but did not use them in the present study. Figure 1 shows examples of recorded stimuli.

Table 1. Speaker 1's sentences average measures

Type	Median Pitch	Hammarberg Index	Valence	Arousal	Duration	Words
Neutral	101.89	21.20	5.42	3.56	1.56	4.93
Prosody	175.41	7.74	5.50	3.80	1.57	4.24
Semantic	102.29	21.27	2.75	6.27	1.55	4.43

Note: All measures are raw averages, acoustic measures are in Hz and duration is in seconds

Table 2. Speaker 2's sentences average measures

Type	Median Pitch	Hammarberg Index	Valence	Arousal	Duration	Words
Neutral	89.20	30.68	5.79	3.67	0.73	2.55
Prosody	129.43	11.36	5.76	4.07	0.93	2.43
Semantic	86.36	28.56	3.05	5.08	0.72	2.40

Note: All measures are raw averages, acoustic measures are in Hz and duration is in seconds

Figure 1. Example of four sentences used in each study. Left panel: sentences recorded by the first speaker. Right panel: sentences recorded by the second speaker. Each panel shows oscillograms and spectrograms per category, green dots in spectrograms indicate fundamental frequency (F0) contours.

Procedure

Rating sessions were conducted online in web-based experiments using the platform Gorilla (gorilla.sc). For session 1 (Speaker 1) participants were recruited via UCL's course credit platform (uclpsychology.sona-systems.com) where they received the Gorilla link for performing the task. For session 2 (Speaker 2) voluntary participants were sent the Gorilla link directly. In both rating sessions participants listened to threatening sentences (Prosody or Semantic) and neutral sentences (Neutral), presented randomly. Participants rated sentences in a 9-point Likert scale for threat (not threatening at all to very threatening). In addition, participants judged sentences arousal and valence levels in a SAM scale (as in ANEW, Warriner et al., 2013). Finally, participants completed a classification task, where they categorised sentences as Calm, Neutral, Angry or Enraged. Half the participants in each session were assigned to judge Prosody and the other half to judge Semantic. For the Prosody task, participants were requested to rate all sentence on how threatening they sounded (obviating their content). For the Semantic task, participants were requested to rate all sentences on how threatening was their content (obviating their sound).

Analysis

To test whether acoustic measures and lexical norms can function as markers for threat, an ordered-logistic regression was implemented to assess the relationship between acoustic/lexical features and threat ratings. Sentence level arousal and valence SAM ratings were assessed in the same manner. The general structure of the model is described on Figure 2. In addition, Bayesian estimation supersede the t-test (BEST) (as in Kruschke, 2013; but using exponential distributions for standard deviations) was used to test differences between measured means of each category. Finally, simple categorical models were used to assess the classification task (Figure 3). All models were sampled with the Hamiltonian Monte Carlo (HMC) no U-turn sampling (NUTS) method as provided by PyMC3 (Salvatier et al., 2016), using 1000 tuning samples and 1000 samples.

Acoustic measures were extracted via Python's Parselmouth interface to Praat (Jadoul et al., 2018). median pitch (fundamental frequency: F0) and Hammarberg index (maximum energy differences between the 0-2000hz and 2000-5000hz ranges). Lexical norms were obtained from the self-assessment manikin (SAM) average ratings on arousal and valence from Warriner et al.'s (2013) study. For both studies, the dependent variable

consisted of threat scores given by participants. Plots and model comparisons were produced using Arviz (Kumar et al., 2019) and Matplotlib (Hunter, 2007).

Figure 2. Graph representation of hierarchical ordered-logistic model, used for analysing ratings data in terms of acoustic measures or lexical norms. The α_s parameter is the re-parametrised varying intercept across subjects. Parameters denoted as β represent slopes of median pitch, Hammarberg index, arousal and valence. Variables represent continuous measures: p (*median pitch*), h (*Hammarberg index*), a (*arousal*), and v (*valence*). Parameter C_n indicates the 9 minus 1 cutpoints for the Likert scale. Note that additional β parameters can be added for shimmer and harmonicity. Observations (obs.) correspond to each single trial rating (800 obs. for Speaker 2's model).

Figure 3. Graph representation of categorical model, used for analysing emotional classification data in terms of acoustic measures or lexical norms. Parameter p is a Dirichlet distribution prior assuming that each *Neutral* or *Prosody* sentence has an equal probability (concentration parameter α) of being categorised as *angry*, *calm*, *enraged* or *neutral*. Observations (obs.) correspond to each single trial where a sentence was classified (800 obs. for Speaker 2's model).

Results

All models showed excellent convergence ($\hat{R} \cong 1$, ESS > 400, and BFMI > 0.7). All other details, such as stimuli, data, summaries, traceplots, autocorrelation-plots and additional plots of results can be also found at this paper's Open Science Framework (OSF) repository (<https://osf.io/xrfkq/>).

Speaker 1 Results

As shown in Table 3, Arousal and Valence contribute little to explain acoustic threat ratings (HDIs across zero), but Median Pitch and Hammarberg Index show clear increases and decreases respectively. This indicates that as pitch increases and voice-quality decreases (more roughness), the ratings become more threatening. Oppositely, Table 4 indicates that pitch and roughness contribute little to explaining semantic threat ratings (HDIs across zero or ROPE), but Arousal and Valence show a strong increase and decrease respectively. Figure 4 summarises cutpoints probabilities.

Table 3. Speaker 1, Prosody experiment, ordered-logistic model estimates

Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	0.033	0.003	0.029	0.037
Hammarberg Index	-0.113	0.013	-0.134	-0.091
Arousal	0.007	0.105	-0.167	0.168
Valence	-0.074	0.158	-0.319	0.199

Note. All estimates are in ordered log-odds.

Table 4. Speaker 1, Semantic experiment, ordered-logistic model estimates

Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	-0.012	0.010	-0.029	0.004
Hammarberg Index	0.028	0.014	0.002	0.050
Arousal	0.518	0.083	0.385	0.653
Valence	0.356	0.056	0.261	0.442

Note. All estimates are in ordered log-odds.

Figure 4. Speaker 1 cutpoints plots. Plots show probability of rating a sentence as not threatening at all (0 points) or very threatening (8 points) given acoustic measures for the Prosody experiment (upper panel) and given lexical norms for the Semantic experiment (lower panel). Prosody orange solid line indicates maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, purple dashed line indicates minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Semantic solid red line indicates maximum Arousal plus minimum valence posterior distributions, blue dashed line indicates minimum Arousal plus maximum Valence posterior distributions. Shadows show 90% HDIs.

For the Prosody experiment, these indicate that when Median Pitch is high and Hammarberg Index is low the probability of ratings below 5 Threat points in the Likert scale is near zero ($\sim 0.2\%$), but considerably increases at 7 points ($\sim 35\%$) and 8 points ($\sim 45\%$). When Median Pitch is low and Hammarberg Index high the probability of rating sentences as zero threatening is near 64% and considerably falls at 2 points ($\sim 17\%$). For the Semantic experiment, when Arousal is high and Valence is low, the probability of rating a sentence as threatening remains below 10% up to 7 points (11%) and jumps up to 64% for 8 points. When Arousal is low and Valence is high the probability of rating zero is near 60% and drops below 10% after 1 point. Note that these are cumulative probabilities, so all scores (9 points, 0 to 8) add up to 1 (100%).

Further confirming these patterns mean differences, estimated via BEST, are remarkably similar. As summarised in Table 5, Arousal means are high for Semantic sentences but low for Neutral and Prosody; while Median Pitch means are high for Prosody sentences but low for Neutral and Semantic. The opposite pattern is observed for Valence and Hammarberg Index. Table 6 summarises estimates effect sizes, and Figure 5 summarises BEST estimates across all conditions. These results indicate that the central tendency (mean) of Median Pitch and Arousal is higher for prosodic and semantic threat respectively, while the central tendency (mean) of Hammarberg Index and Valence is lower for prosodic and semantic threat respectively.

Table 5. Speaker 1, estimates from BEST models

Study 1	Arousal		Valence		Median Pitch		Hammarberg Index	
<i>Category</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Neutral	3.54	0.58	5.43	0.36	101.46	4.14	21.09	3.53
Prosody	3.79	0.51	5.50	0.36	175.75	10.91	7.84	2.83
Semantic	6.26	0.63	2.75	0.85	102.07	4.26	21.22	3.77

Table 6. Speaker 1, mean differences effect sizes (d)

Difference	Arousal	Valence	Median Pitch	Hammarberg Index
Neutral - Prosody d	-0.46	-0.21	-9.10	4.16
Neutral - Semantic d	-4.51	4.15	-0.15	-0.04
Prosody - Semantic d	-4.31	4.27	8.99	-4.03

Figure 5. Speaker 1 differences of means. BEST estimates. Top: Arousal and Valence (difference scores from lexical SAM average values). Bottom: Median Pitch and Hammarberg Index (difference scores in Hz). Circles represent the mean and lines represent 90% HDIs from mean differences' posterior distributions. Grey bands indicate 2SDs regions of practical equivalence (ROPE).

Figure 6. Speaker 1 SAM ratings cutpoints plots. Plots show probability of rating a sentence as low in arousal/valence (0 points) or high in arousal/valence (8 points) given acoustic measures for the Prosody experiment (upper panels) and given lexical norms for the Semantic experiment (lower panels). Upper panels: Orange solid lines indicate maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, purple dashed lines indicate minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Lower panels: Solid red lines indicate maximum Arousal plus minimum valence posterior distributions; blue dashed lines indicate minimum Arousal plus maximum Valence posterior distributions. Shadows show 90% HDIs.

Speaker 1's SAM ratings for arousal and valence given to each sentence provide further confirmation of this pattern. These are not the original ratings from Warriner et al (2013), but ratings given by present study's participants to sentences containing those words; they can thus be denominated sentence-arousal and sentence-valence. Figure 6 summarises the posterior distributions of sentence SAM ratings when sentences' acoustic measures and lexical norms are at their maximums or minimums. For the Prosody experiment this indicates that for max median pitch (MP) and min Hammarberg index (HI), the probability of giving 8 sentence-arousal points is near 64% and of giving 7 points is near 27%. For min MP and max HI, the probability of giving 0 points is near 47% and of giving 1 point is near 20%. Differently for max MP and min HI sentence-valence scores of 0 points are near 23% and for 1 point near 39%. For min MP and max HI they concentrate around 4 points (the median) with around 35% probability. These results indicate that high MP and low HI predict high arousal and low valence SAM ratings at the sentence level. The Semantic experiment (Figure 6 lower panels), shows a very similar pattern, indicating that lexical items' high arousal and low valence predict high arousal and low valence SAM ratings at the sentence level.

Average ratings for Prosody are indicative of the same pattern, high sentence-arousal ($m = 5.96$, $SD = 1.4$) and low sentence-valence ($m = 2.47$, $SD = 1.49$) for Prosody sentences (prosodic threat only), as opposed to Neutral sentences showing low sentence-arousal ($m = 1.64$, $SD = 1.66$) and sentence-valence around the median ($m = 4.25$, $SD = 1.86$). Average ratings for Semantic indicate high sentence-arousal ($m = 6.08$, $SD = 2.10$) and low sentence-valence ($m = 1.65$, $SD = 1.52$) for Semantic sentences (semantic threat only), but low sentence-arousal ($m = 2.24$, $SD = 2.01$) and sentence-valence around the median ($m = 4.19$, $SD = 1.19$) for Neutral sentences. Finally, Table 7 summarises the categorical model estimates. In the Prosody experiment, ~88% of the time Prosody sentences (threatening sound) are classified as angry-related emotions (~52% angry and ~28% enraged), while Neutral sentences tend to be classified as neutral (~64%) or calm (~33%). In the Semantic experiment, ~80% of the time Semantic sentences (threatening content) are classified as anger-related (~52% angry and ~28% enraged), while Neutral sentences tend to be classified as neutral (~60%) or calm (~32%). These results indicate when participants focus either on sound or content, they tend to classify the speaker as being angry for threatening sentences and neutral/calm for non-threatening sentences.

Table 7. Speaker 1, Categorical Model Estimates

<i>Prosody Experiment</i>					<i>Semantic Experiment</i>				
Stimulus	Response	Mean	HDI 5%	HDI 95%	Stimulus	Response	Mean	HDI 5%	HDI 95%
Neutral	Angry	3.17	2.06	4.45	Neutral	Angry	6.16	4.48	7.78
Neutral	Calm	32.57	29.37	35.97	Neutral	Calm	31.87	28.54	34.84
Neutral	Enraged	0.79	0.20	1.34	Neutral	Enraged	2.27	1.21	3.31
Neutral	Neutral	63.48	60.10	67.10	Neutral	Neutral	59.69	56.19	62.88
Prosody	Angry	72.56	69.47	75.92	Semantic	Angry	51.81	48.28	55.08
Prosody	Calm	0.41	0.05	0.79	Semantic	Calm	3.53	2.30	4.86
Prosody	Enraged	15.78	13.26	18.48	Semantic	Enraged	27.45	24.27	30.58
Prosody	Neutral	11.25	9.10	13.42	Semantic	Neutral	17.21	14.51	19.77

Note. All estimates are percentages. Means represent the estimated proportion of response given a stimulus type.

Figure 7 summarises predictions of Speaker 2's ratings based on Speaker 1's ratings. Speaker 1's voice has (on average) a pitch around 30Hz higher than Speaker 2's voice. This represents a nearly 30% higher pitch than Speaker 2's average recorded pitch (109Hz). So, Speaker 1's Median Pitch was scaled down by 30%. This pitch differences are also reflected in Speaker 2's Hammarberg Index, which increases around 20% on average. Hence, Speaker 1's Hammarberg Index was scaled up by 20%. These scalings are akin to a baseline correction and allow to account for error in lack of specific prior assumptions about speakers' pitch differences. The average prediction error (difference between observed and predicted scores) for the Prosody experiment was of 0.32 points for Neutral and of 0.58 points for Prosody. The average prediction error (difference between observed and predicted scores) for the Semantic experiment was of 1.01 points for Neutral and of 1.25 points for Semantic. Overall, this implies an excellent predictive accuracy for Prosody and good predictive accuracy for Semantic across the two speakers despite the substantial differences in sentence contents.

Figure 7. Prediction of ratings for Speaker 2 using ratings from Speaker 1. Upper left: Prosody predictions across Median Pitch (pitch). Upper right: Prosody threat predictions across Hammarberg Index (roughness). Bottom left: Semantic threat predictions across Arousal. Bottom right: Semantic threat predictions across Valence. Note that Semantic predictions are less accurate for threat and Prosody predictions are less accurate for neutral.

Speaker 2 Results

Results from the models (same as for Speaker 1) applied to Speaker 2's data, indicate the same patterns. Table 8 summarises slopes from the Prosody experiment, these indicate a clear increase in Median Pitch and decrease in Hammarberg index, while Valence slightly decreases but overlaps ROPE, Arousal seem to have a minor effect. This effect can be explained by the difficulty of finding truly semantically neutral words that fit the restrictive sentence context ("You're (a/an) ____"), which might have resulted in a somewhat unbalanced sample, where Neutral sentences got words with higher arousal scores than Prosody. Nevertheless, it is clear that acoustic features (Median Pitch and Hammarberg Index) drive the effect. Table 9 summarises slopes from the Semantic experiments. These indicate strong effects of Arousal and Valence, while Median Pitch falls across zero and Hammarberg Index shows a small effect but overlaps ROPE.

Table 8. Speaker 2, Prosody experiment, ordered-logistic model estimates

Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	0.050	0.006	0.040	0.060
Hammarberg Index	-0.093	0.012	-0.113	-0.073
Arousal	-0.346	0.079	-0.469	-0.218
Valence	-0.158	0.086	-0.310	-0.026

Note. All estimates are in ordered log-odds.

Table 9. Speaker 2, Semantic experiment, ordered-logistic model estimates

Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	0.000	0.022	-0.033	0.038
Hammarberg Index	-0.030	0.011	-0.049	-0.012
Arousal	0.266	0.085	0.129	0.407
Valence	-0.836	0.066	-0.936	-0.723

Note. All estimates are in ordered log-odds.

Confirming this pattern of results, Figure 8 summarises rating probabilities (cutpoints by measures). For the Prosody experiment, these indicate that when Median Pitch is high and Hammarberg Index is low the probability of ratings below 6 Threat points in the Likert scale is near zero (up to ~2.6%), increases up to 10% for 7 and 8 points, but considerably increases at 9 points (~74%). When Median Pitch is low and Hammarberg Index high the probability of rating sentences as 1 is around 55% and considerably falls at 2 points (~23%) and below 10% from 3 points. For the Semantic experiment, when Arousal is high and Valence is low, the probability of rating a sentence as threatening remains below 10% up to 7 and 8 points (both ~16%) and jumps up to near 40% for 8 points. When Arousal is low and Valence is high the probability of rating 1 is near 62%, drops to near 13% for 2 points and below 10% afterwards. Note that these are cumulative probabilities, so all scores (9 points, 0 to 8) add up to 1 (100%).

Table 10. Speaker 2, estimates from BEST models

Study 1	Arousal		Valence		Median Pitch		Hammarberg Index	
<i>Category</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Neutral	3.70	0.82	5.92	0.89	89.21	2.22	31.26	4.81
Prosody	4.14	0.99	6.01	0.99	129.67	7.75	11.06	2.99
Semantic	4.83	0.75	3.09	0.83	86.50	2.56	28.51	5.66

Figure 8. Speaker 2 cutpoints plots. Plots show probability of rating a sentence as not threatening at all (1 points) or very threatening (9 points) given acoustic measures for the Prosody experiment (upper panel) and given lexical norms for the Semantic experiment (lower panel). Prosody orange solid line indicates maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, purple dashed line indicates minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Semantic solid red line indicates maximum Arousal plus minimum valence posterior distributions, blue dashed line indicates minimum Arousal plus maximum Valence posterior distributions. Shadows show 90% HDIs.

Figure 9. Speaker 2 differences of means. BEST estimates. Top: Arousal and Valence (difference scores from lexical SAM average values). Bottom: Median Pitch and Hammarberg Index (difference scores in Hz). Circles represent the mean and lines represent 90% HDIs from mean differences' posterior distributions. Grey bands indicate 2SDs regions of practical equivalence (ROPE).

Table 11. Speaker 1, mean differences effect sizes (d)

Difference	Arousal	Valence	Median Pitch	Hammarberg Index
Neutral - Prosody d	-0.49	-0.10	-7.18	5.09
Neutral - Semantic d	-1.45	3.31	1.14	0.53
Prosody - Semantic d	-0.80	3.23	7.57	-3.89

Again, BEST results confirm this pattern. As summarised in Table 10, Arousal means are high for Semantic respect to Neutral and Prosody; while Median Pitch means are higher for Prosody respect to Neutral and Semantic. Again, the opposite pattern is observed for Valence and Hammarberg Index. Figure 9 summarises the differences between means of sentences recorded by Speaker 2, as estimated by BEST models, and Table 11 summarises the effect size of each difference between means.

Figure 10. Speaker 2 SAM ratings cutpoints plots. Plots show probability of rating a sentence as low in arousal/valence (0 points) or high in arousal/valence (8 points) given acoustic measures for the Prosody experiment (upper panels) and given lexical norms for the Semantic experiment (lower panels). Upper panels: Orange solid lines indicate maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, purple dashed lines indicate minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Lower panels: Solid red lines indicate maximum Arousal plus minimum valence posterior distributions; blue dashed lines indicate minimum Arousal plus maximum Valence posterior distributions. Shadows show 90% HDIs.

Speaker 2's SAM ratings provide further confirmation of this pattern. Figure 10 summarises the posterior distributions of sentence SAM ratings when sentences' acoustic measures and lexical norms are at their maximums or minimums. For the Prosody experiment this indicates that for max median pitch (MP) and min Hammarberg index (HI), the probability of giving 9 sentence-arousal points is near 69% and of giving 8 points is near 18%. For min MP and max HI, the probability of giving 1 point is near 27%, around 15% for 2 points and near 19% for 3 points. Differently for max MP and min HI sentence-valence scores of 1 are near 43%, around 21%, and near 16% for 3 points. For min MP and max HI they concentrate around 5 points (the median) with around 38% probability. These results indicate that high MP and low HI predict high arousal and low valence SAM ratings at the sentence level. The Semantic experiment (Figure 10 lower panels), shows

a very similar pattern, indicating that lexical items' high arousal and low valence predict high arousal and low valence SAM ratings at the sentence level.

Table 12. Speaker 2, Categorical Model Estimates

<i>Prosody Experiment</i>					<i>Semantic Experiment</i>				
Stimulus	Response	Mean	HDI 5%	HDI 95%	Stimulus	Response	Mean	HDI 5%	HDI 95%
Neutral	Angry	3.46	0.93	2.02	Neutral	Angry	3.46	0.89	1.97
Neutral	Calm	54.76	2.38	50.59	Neutral	Calm	24.69	2.13	21.05
Neutral	Enraged	4.46	1.01	2.77	Neutral	Enraged	0.74	0.43	0.12
Neutral	Neutral	37.32	2.36	33.36	Neutral	Neutral	71.10	2.31	67.10
Prosody	Angry	51.79	2.49	47.88	Semantic	Angry	55.77	2.36	52.16
Prosody	Calm	4.93	1.05	3.24	Semantic	Calm	1.96	0.71	0.79
Prosody	Enraged	16.64	1.87	13.87	Semantic	Enraged	15.08	1.80	12.23
Prosody	Neutral	26.64	2.18	23.25	Semantic	Neutral	27.19	2.18	23.48

Note. All estimates are percentages. Means represent the estimated proportion of response given a stimulus type.

Table 12 summarises the categorical model estimates. In the Prosody experiment, ~69% of the time Prosody sentences (threatening sound) are classified as anger-related (~52% angry and ~17% enraged), while Neutral sentences tend to be classified as neutral (~37%) or calm (~55%). In the Semantic experiment, ~71% of the time Semantic sentences (threatening content) are classified as anger-related (~56% angry and ~15% enraged), while Neutral sentences tend to be classified as neutral (~71%) or calm (~25%). These results indicate when participants focus either on sound or content, they tend to classify the speaker as being angry for threatening sentences of the appropriate type and neutral/calm for the others (e.g. appropriately indicating in the Prosodic experiment that Semantically threatening sentences spoken in neutral prosody did not “sound threatening”). Note in the Prosody experiment, the higher percentage for calm and slightly lower for anger may be explained by the generally lower pitched voice of Speaker 2.

Overall, Speaker 2 results replicate Speaker 1 results, as expected by the precise predictions (Figure 6). This is despite that the latter has two items with positive/neutral value on their own but with very negative value given sentence context: ‘animal’ (arousal = 4.3, valence = 7.06) in ‘You’re an animal’ (arousal-sen = 7.1, valence-sen = 2.1), and ‘pig’ (arousal = 3.68, valence = 4.83) in ‘You’re a pig’ (arousal-sen = 6.8, valence-sen = 2). This relates to the fact that, as mentioned above, Arousal tends to be slightly higher than

expected for Prosody sentences. This is reflected by Arousal differences between means in Figure 7, which show weaker effects but still barely overlap ROPE or other posteriors distributions. Even so, Speaker 2's SAM ratings for arousal (sentence-arousal) and valence (sentence-valence) given to each sentence are clearly similar to Speaker 1's results, with the sole difference of a generally higher valence for Speaker 2.

Discussion

Present results show that median pitch and Hammarberg index are sufficient for predicting threatening speech comprehension. In addition, when prosody is absent from sentences, but sentences contain threatening (offensive, harmful) lexical items, arousal and valence norms provide analogous results respect to prosody. In other words, a voice that is more pitched and with less quality (rougher) is understood as threatening and angry by participants, and sentences containing words high in arousal and low valence are also understood as threatening and angry. This is supported by clear mean differences between acoustic parameters by category, by higher probabilities of rating sentences as more threatening when median pitch/arousal are high and Hammarberg index/valence are low, and by higher rates of classifying these sentences as angry. Finally, these results are consistent across the two speakers, which is shown by good predictive capacity of models and by a replicated pattern of results from one speaker to another.

Therefore, present results add up to the accumulated evidence indicating that more pitch and roughness are expressions of hot-anger (Banse and Scherer, 1996; Frühholz et al., 2016; Hammerschmidt and Jürgens, 2007; Juslin and Laukka, 2003). Although most of these previous investigations usually use harmonicity (harmonic to noise ratio) as a measure of noisier/lower quality speech, present results show that Hammarberg index, a frequency measure of voice-quality, can be a better predictor. Indeed, angry prosody's harmonicity has not always shown direct effects on brain activity (Frühholz et al., 2016). Importantly, noting that most Prosody sentences are classified as expressing anger or rage, the present approach links this hot-anger/rage production with aggression, namely with the communication of threat. This implies that the speakers in the present study, despite their lack of training in acting or voice acting, effectively communicated threatening speech (as we requested from them). This is consistent with research observing that natural expressions of anger have increased pitch and decreased

voice-quality (Anikin and Lima, 2017). This implies that sentences with more pitch and rougher (less voice-quality) are naturally comprehended as threatening.

In terms of semantics, present results support the sufficiency of Valence and Arousal norms (Warriner et al., 2013), at least for the portrayal of semantic threat. This is supported from previous evidence showing how pain or threat corresponds to increases in negativity and arousal (Borelli et al., 2018; Ho et al., 2015). This indicates that these words convey the potential of inducing harm (indirect or potential threat), which is consistent with prosody measures as conveying potential aggression by increasing pitch but decreasing voice quality. In addition, present results indicate that lexical items can strongly drive the interpretation of threatening sentences. This is consistent with previous evidence evaluating semantic emotional effects of sentences based on emotional words (e.g. Kotz and Paulmann, 2007; Chen et al., 2011). However, syntax may also drive the perception or comprehension of sentences as threatening. Indeed, it has been shown that sentences lacking lexical items with emotional meaning can be comprehended as emotional (Lai et al., 2015). So, even if present lexical items are sufficient for explaining and predicting perceived threat, addressing possible influences of syntax can prove a relevant further step.

Present acoustic measures and lexical norms are sufficient for inferring and predicting threat comprehension, but they cannot provide a full account of the prosody and semantics of anger, which may be composed by a myriad of features establishing more complex and nuanced relations (Lausen and Hammerschmidt, 2020). Having stated this, the present analogy between bidimensional features of prosody and semantics is further confirmed by SAM and classification tasks. SAM ratings are clear in this respect, with strong probability increases for high arousal and low valence ratings when sentences are prosodically or semantically threatening. And when participants are asked how the speaker sounds (prosody) or what he means (semantic), they classify prosodic/semantic sentences as expressing anger in a bigger proportion or rage (in a minor proportion). Taken together, these tasks demonstrate that participants do not only understand sentences as threatening, but also as more arousing and negative (low in valence), and as expressing anger. Furthermore, all tasks demonstrate that Neutral sentences are clearly distinguished from Semantic and Prosody sentences in terms of threat, high arousal, low valence, and anger/rage.

All in all, our study provides evidence for the efficiency of a more minimalistic approach to the characterisation of threatening prosody. Neuroscientific theories emphasise the role of spectral features of prosody, as these reflect anatomical (i.e. laterality, dorsality) and temporal properties of speech processing (Poeppel et al., 2008; Zatorre et al., 2002). Based on these, we proposed that a bidimensional set of features, where the increase of a feature (pitch) and the decrease of another (voice-quality) would be sufficient for inferring and predicting the comprehension of speech stimuli as prosodically threatening. Furthermore, we proposed that it could be possible to find analogous features in the semantic domain, namely increasing arousal and decreasing valence. Emphasising that present hypotheses propose these analogous features as sufficient but not exhaustive, our observations confirm these hypotheses and show that the present approach is a step forward in researching naturalistic speech. Besides implementing more transparent and appropriate statistical modelling and stimuli selection (Alday, 2018), the present approach is a strong alternative to conventional methods for treating experimental stimuli, usually subject to inappropriate procedures and statistical treatment (Sassenhagen and Alday, 2016). Also, present results demonstrate the appropriateness of Bayesian statistical approaches to small data problems, allowing for better characterisation of error, precision, and uncertainty, which is difficult to attain in more conventional frequentist approaches to emotional characterisation across multiple emotional types and measures (e.g. Lausen and Hammerschmidt, 2020). This does not imply that the present approach is necessarily better, but that it is a good alternative, in particular when working on a relatively small scale.

Besides the potential contributions of the present study, there are some limitations that need to be addressed. Firstly, individual characteristics of voices of different speakers can impact the predictive capacity of models, and we addressed this issue in a rudimentary manner (by proportionally decreasing pitch of Speaker1 respect to Speaker 2 for making predictions). In future research, this could be addressed by directly including error parameters related with inter-speaker voice variation in the models, as well as including much wider range of speaker demographics. Secondly, although more natural sentences (Speaker 1) were compared to more controlled ones (Speaker 2), the core of our recording session is still in a highly controlled lab set up. Further research could be conducted in fully natural environments to test whether

present hypotheses hold for situations of more natural threatening speech production. Despite this, the consistency of predictions and inference across different types of sentences and speakers evidences that Bayesian approaches are an excellent option for dealing with small data, and also demonstrates that hierarchical models can properly address data inconsistencies if appropriate and meaningful levels and distributions (i.e. ordered-logistic for rating scales) are included in models. Also, the present approach may prove useful for future practical applications requiring predictive capacity which is not theoretically/inferentially detached.

In summary, the present study shows strong acoustic and lexical counterparts that underlay prosodic and semantic threat. Namely, Arousal increases and Valence decreases, indicating that semantic threat is comprehended via negatively arousing content. Similarly, Median Pitch increases and Hammarberg Index decreases suggest that prosodic threat is comprehended through a rougher higher pitched voice. Noting as well that variation in these features also imply that present stimuli are comprehended as expressing anger. This comparability between semantic and prosody features supports an analogous treatment of threat comprehension via minimal content and acoustic features, which allow for strong inference and predictions. Present evidence extends the scope of evidence supporting theories of threat production and perception and of accumulated evidence regarding features of angry speech.

Authors Contributions

Contributions to the study were as follows. **Busch-Moreno**: Conceptualization, Methodology, Software, Analysis, Investigation, Resources, Curation, Writing, Visualization, Administration. **Vinson**: Conceptualization, Methodology, Revision, Supervision, Administration.

Data statement

The data that support the findings of this study are openly available in the Open Science Framework (osf.io) at <https://osf.io/xrfkq/> [DOI 10.17605/OSF.IO/XRFBKQ].

Declaration of interests

The authors report no conflict of interest.

Funding Sources

This work was supported by CONICYT, Becas Chile, under Grant number 72170145.

References

- Anikin, A., & Lima, C. F. (2017). Perceptual and Acoustic Differences between Authentic and Acted Nonverbal Emotional Vocalizations. *Quarterly Journal of Experimental Psychology*, 71(3), 17470218.2016.1.
<https://doi.org/10.1080/17470218.2016.1270976>
- Alday, P. M. (2018). M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*, 34(4), 457–473.
<https://doi.org/10.1080/23273798.2018.1546882>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
<https://doi.org/10.1037/0022-3514.70.3.614>
- Borelli, E., Crepaldi, D., Porro, C. A., & Cacciari, C. (2018). The psycholinguistic and affective structure of words conveying pain. *PLOS ONE*, 13(6), e0199658.
<https://doi.org/10.1371/journal.pone.0199658>

- Busch-Moreno, S., Tuomainen, J., & Vinson, D. (2021). *Trait Anxiety Effects on Late Phase Threatening Speech Processing: Evidence from Electroencephalography*. In press.
- Burra, N., Kerzel, D., Munoz Tord, D., Grandjean, D., & Ceravolo, L. (2018). Early spatial attention deployment toward and away from aggressive voices. *Social Cognitive And Affective Neuroscience*, 14(1). <https://doi.org/10.1093/scan/nsy100>
- Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, 86(3), 158–167. <https://doi.org/10.1016/j.biopsycho.2010.11.004>
- Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: Time course and ERP correlates. *Neuroscience Letters*, 533(533), 90–95. <https://doi.org/10.1016/j.neulet.2012.10.054>
- Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive, Affective, & Behavioral Neuroscience*, 16(3), 415–432. <https://doi.org/10.3758/s13415-016-0402-y>
- Frühholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A.-K., Vuilleumier, P., & Grandjean, D. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective valence. *Social Cognitive and Affective Neuroscience*, 11(10), 1638–1649. <https://doi.org/10.1093/scan/nsw066>
- Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, 21(5), 531–540. <https://doi.org/10.1016/j.jvoice.2006.03.002>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Ho, S. M. Y., Mak, C. W. Y., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional Valence, Arousal, and Threat Ratings of 160 Chinese Words among Adolescents. *PLOS ONE*, 10(7), e0132294. <https://doi.org/10.1371/journal.pone.0132294>
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/mcse.2007.55>
- Imbir, K. K., Spustek, T., Duda, J., Bernatowicz, G., & Żygierewicz, J. (2017). N450 and LPC Event-Related Potential Correlates of an Emotional Stroop Task with Words Differing in Valence and Emotional Origin. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00880>

- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15.
<https://doi.org/10.1016/j.wocn.2018.07.001>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kemmerer, D. L. (2015). *Cognitive neuroscience of language*. Psychology Press.
- Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151(50), 107–118.
<https://doi.org/10.1016/j.brainres.2007.03.015>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573–603.
<https://doi.org/10.1037/a0029146>
- Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, 4(33), 1143. <https://doi.org/10.21105/joss.01143>
- Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the Lines: Implied Emotion in Sentence Comprehension. *Journal of Cognitive Neuroscience*, 27(8), 1528–1541. https://doi.org/10.1162/jocn_a_00798
- Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, 7(1). <https://doi.org/10.1057/s41599-020-0499-z>
- Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, 105(1), 59–69.
<https://doi.org/10.1016/j.bandl.2007.11.005>
- Pell, M.D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, 111, 14–25.
<https://doi.org/10.1016/j.biopsycho.2015.08.008>
- Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1071–1086.

- <https://doi.org/10.1098/rstb.2007.2160>
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2, e55. <https://doi.org/10.7717/peerj-cs.55>
- Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, 162, 42–45. <https://doi.org/10.1016/j.bandl.2016.08.001>
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- Yan, M., & Sommer, W. (2018). The effects of emotional significance of foveal words on the parafoveal processing of N + 2 words in reading Chinese sentences. *Reading and Writing*, 32(5), 1243–1256. <https://doi.org/10.1007/s11145-018-9914-x>
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46. [https://doi.org/10.1016/s1364-6613\(00\)01816-7](https://doi.org/10.1016/s1364-6613(00)01816-7)
- Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H., & Luo, Y. (2014). Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience*, 9(12), 1897–1903. <https://doi.org/10.1093/scan/nst188>