ELS Journal on Interdisciplinary Studies in Humanities
Volume 4 Issue 1, 2021
DOI: https://doi.org/10.34050/elsjish.v4i1.12786
Homepage: journal.unhas.ac.id/index.php/jish

# Facial Expressions Alter the Fundamental Sound Properties of Speech

Lucas W.E. Tessaro[1*], Cynthia Whissell[2]

[1]Behavioural Neuroscience Program, Department of Psychology, Interdisciplinary Human Studies, Laurentian University, Canada
[2]Department of Psychology, Interdisciplinary Human Studies, Laurentian University, Canada
*Correspondence: lx_tessaro@laurentian.ca

**ABSTRACT**

*Literature from across academic disciplines has demonstrated significant links between emotional valence and language. For example, Whissell's Dictionary of Affect in Language defines three dimensions upon which the emotionality of words is describable, and Ekman's Theories of Emotion include the perception and internalization of facial expressions. The present study seeks to expand upon these works by exploring whether holding facial expressions alters the fundamental speech properties of spoken language. Nineteen (19) participants were seated in a soundproof chamber and were asked to speak a series of pseudowords containing target phonemes. The participants spoke the pseudowords either holding no facial expression, smiling, or frowning, and the utterances recorded using a high-definition microphone and phonologically analyzed using PRAAT analysis software. Analyses revealed a pervasive gender differences in frequency variables, where males showed lower fundamental but higher formant frequencies compared to females. Significant main effects were found within the fundamental and formant frequencies, but no effects were discerned for the intensity variable. While intricate, these results are indicative of an interaction between the activity of facial musculature when reflecting emotional valence and the sound properties of speech uttered simultaneously.*

**ARTICLE HISTORY**
Published March 27th 2021

Check for updates

**KEYWORDS**
PRAAT; Phonological; Phoneme; Gender Differences; Mean Pitch



## 1. Introduction

Whether phonemes carry meaning in and of themselves or whether one ascribes meaning arbitrarily or through convention has been the subject of phonosymbologist discourse across disciplines for the better part of the last 2500 years. The discussion itself could be said to originate at least with Plato in the Socratic Dialogue *Cratylus*, in which Socrates eventually reveals his (Plato's) position to be a synergism of the two – some sounds may be naturally expressive while others are completely arbitrary (Plato, 2015). Many esteemed scholars of Aristotle point out his desire to seek a compromise between the two: although the relationship between written and spoken words is conventional, as is the mental state evoked by words and their intended meaning, the relation between mental state and external object is universal – thus, different languages use different sounds for the same object (Ackrill, 1975; Modrak, 2001). Aristotle's approach and emphasis on the purpose and use of the definition indicate the ostensive relation between metaphysical truth assertions and extralinguistic objects; that is, a heavy component of Aristotle's philosophy and metaphysics depends on the assumption that linguistic statements are true if and only if they correspond to some reality (Ackrill, 1975; Modrak, 2001).

An extensive body of literature exists bridging the gap between Aristotle and Wittgenstein, however, it is within the mind of the 20th Century philosopher that some of the most important conceptualizations of language occurred; in stark contrast to Aristotle's realism and truth, Wittgenstein sees only unverifiable subjectivity. Wittgenstein posits that language is more akin to a game that we all know how to play; it is a strongly metaphorical game, and he uses it to illustrate the fact that many games proceed *without* conscious reference or adherence to the rules – we just play the game[1]. Yet within this game, any definition of a word presents itself with a plausible counterexample, rendering any attempts to completely define it impossible – which would be in direct opposition to Aristotle (Ackrill, 1975; Modrak, 2001; Wittgenstein, 2010). Despite

[1] The equating of language interactions with games should not be taken too literally

this inability to come to universal definitions, we as a species all know how to use language, and through additional situational context and use we are able to discern what words mean, even potentially meaningless ones or those used inappropriately (Wittgenstein, 2010).

Wittgenstein argued that the connection made between words and meanings occurs through family resemblances of clustered concepts - if at no point did **TABLE** refer to the overarching concept of a **TABLE**, then the word would be as utterly meaningless as *gruklit*. His approach is fairly amendable to modern psychological models of language and learning, reflecting the basic processes of pair associating and generalizing, and to some extent reflect neuroscientific analyses of processing linguistic information and the localization of semantic knowledge (Fine, 2008; Grisoni, Miller, & Pulvermüller, 2017; Huth, De Heer, Griffiths, Theunissen, & Gallant, 2016b; Huth, Nishimoto, Vu, & Gallant, 2012; Pestilli, 2018). Wittgenstein might explain this in terms of the original **TABLE** being the **PARADIGM CASE** for any table, and as the word becomes used in more abstract ways (*e.g.* a table of contents*)*, it becomes a **FRINGE CASE**, less resembling of the original meaning intended by **TABLE**. Here we might consider Frege's distinction between the **SENSE** a word evokes – its family resemblances of clustered concepts – and the physical object it explains, its **REFERENT**, in order for meaning to be possible (Frege, 1948). A similar discussion occurs within Julian Jayne's *Breakdown of the Bicameral Mind*, wherein he uses the terms definer, definand, metaphier, and metaphrand in much the same way; essentially, words obtain meaning through their close resemblance to some reality under attempt of verbal description, much as stated above (Jaynes, 2000). A thorough analysis is beyond this paper, but suffice it to say Wittgenstein's postulations were not *ex nihilo*, and his emphasis on the subjective experience of language is paramount.

Of the various linguists and psychologists who have directed the discussion on phonosymbolism, de Saussure, Skinner and Chomsky have been amongst the most influential. De Saussure argued strongly in favor of the arbitrariness of the sign, that any word can represent any concept so long as there is consensus about it (De Saussure, 2011). Moreover, he stressed that the two parts of a word, the **SIGNIFIER** (the sign itself) and the **SIGNIFIED** (the thing represented by the sign), are inseparable and impossible to conceptualize as anything other than a single entity; these definitions likely served as the original basis for the terms devised by Jaynes above (Jaynes, 2000). Skinner viewed language as any other act humans engage in, a behaviour that can be explained in terms of *stimulus, response,* and *reinforcement* (Skinner, 2014). Skinner's work has received a significant amount of criticism, including the one that it is almost entirely a theoretical treatise and has large conceptual problems, while many of de Saussure's initial theories have expanded and improved their accuracy in detailing human linguistic behaviour. In his most critical review, Chomsky dismisses much of Skinner's emphasis on the influence of the environment and setting during which the learning of a word occurs (stimulus/response pairing) as a gross oversimplification of a complex operation (Noam Chomsky, 1959). Interestingly, Chomsky's work generated advances in linguistics still seen as fundamental to the field, as well as being key to the development of the interdiscipline of cognitive science, whereas those who follow Skinner's strict behavioural approaches are a select few.

Perhaps the most salient content in which we should explore the subject of phonosymbolism is within the context of human emotion. Knowledge of the internal mental state of another, central if not the key feature of **THEORY OF MIND**, is most frequently assessed in human interaction with probing questions such as "How are you?" (Leslie, 1987) Complex social situations notwithstanding, how one makes the interrogative and the method of response in this particular human interaction forms the basis and foundation of almost any other to follow. The importance of understanding how humans express their emotions through such vocalizations is of the utmost importance – emotion features significantly in our species' social interactions (Feld, 1981). That the words themselves might encode some aspect of emotional information evidences from studies of primate vocalizations. When viewed in the context of early primates, similar vocalizations expresses multiple affective states by changing their phonological properties of the vocalizations (Chevalier-Skolnikoff, 1973; Sauter, Eisner, Ekman, & Scott, 2010). Primate calls are not stereotyped, but rather fall under a common *pattern* and the phonological properties of a pattern are altered to infer complex, subtle meaning as well as giving notice of an affective state (Scherer, 1985).

When an individual experiences an emotion, there is a physiological response, one which influences all biological systems including those involved with speech production (Simonyan, Ackermann, Chang, & Greenlee, 2016). Conceptually, the increased tension throughout the body one experiences when sobbing also occurs within the vocal chords, potentially increasing pitch; an increase in heart rate and the flush one feels when angry might reflect a change an intensity, creating more energy per unit sound. Strong links between the physical properties of speech and underlying emotional affect have been noted in previous studies (Murray & Arnott, 1993). It should thus not be surprising that recent works have further demonstrated this link between the phonological properties of sounds and a related emotional affective state. It is in this vein of research that Whissell constructed the **DICTIONARY OF AFFECT IN LANGUAGE**, or DAL, following a tradition of lexical-emotional analysis (Whissell, 1989, 2009). In several works since its inception, Whissell has demonstrated the validity of the DAL in ascertaining the underlying emotional meaning in written text along three affective dimensions: Pleasantness, Activation, and Imagination (Whissell, 2000, 2001, 2006).

A small gap in the literature exists stemming from Whissell's work as to what extent humans modulate this underlying emotional content in verbal interaction, if consciously at all, and similarly how an interlocutor responds to these emotional modulations; if anything, the limiting step of the DAL is its reliance on written reports and assessments[2] (Whissell, 2009). The overall gap this work seeks to fill is the connection between words observed and their assessments of emotionality given; if a word is pleasant, but I am unhappy, what is the overall outcome of emotional affect? The connective structure, as discussed, is the vocal apparatus; it would be expected that altering the structure of the vocal apparatus would similarly change its function – modulating the sound produced (Simonyan et al., 2016). However, it is questionable what role the orofacial cavity may have in this modulation in respect to emotional affect. The mouth and lips, subtleties such as differences in vowels and consonants also code many of the finer aspects of language. Yet facial expressions and human emotions are intimately intertwined, often occurring as constants across cultures as aptly described by Ekman (Ekman & Friesen, 1971; Ekman & Oster, 1979). Taking the works of Whissell and Ekman together, the present study seeks to explore whether facial expressions can alter the phonological properties of speech sounds. In this way, we hope to enquire to what extent Ekman's emotional theories are synergized with Whissell's DAL when investigated in the context of spoken rather than written language.

## 2. Method

### 2.1 Participants

Prior to recruitment the experimental design received ethics approval from the Laurentian University Research Ethics Board (LU REB). After approval, participants we recruited from Undergraduate Psychology courses where the principal investigator presented a general overview of the experiment and asked those interested to leave their university email addresses to arrange a time to enter the lab. We applied no exclusionary criteria to the 24 participants who volunteered, though lost four due to attrition. We removed one final participant from the database as the audio recordings from this participants' session were not useable. The findings here are thus based upon 19 participants (8 male; 11 female) aged 20-30 (median 22). Recordings were conducted while the participant was comfortably seated within a soundproof room with the experimenter.

### 2.2 Experimental Design and Stimulus

Participants were seated in a comfortable chair enclosed in a soundproof chamber with warm ambient lighting. The recording microphone and experimenter's laptop was on a small desk in front of the participant. After signing consent forms, participants were briefed on the actual study – they would receive a piece of paper upon with several pseudowords; contained in each of those pseudowords were target vowels and the aim of the study was to observe how facial expressions held when saying the pseudowords altered the physical properties of the sounds. The participants had the opportunity to view the words beforehand to clarify expected pronunciation, and then the experiment began. We then asked the participants to read the line of pseudowords corresponding to the number read by the experimenter naturally as if reciting for someone new to the language (not slowly or artificially, but with a natural prosody and enunciation). After the participant read the entire sheet of pseudowords, a brief 2-minute pause was given and then the participant was prompted to read the list once more, this time holding a *smiling* facial expression. After this session of recordings, the list was read a third time following a short 2-minute pause with the *frowning* facial expression. We recorded and saved each line individually using Audacity® v2.1.2 open source, cross-platform audio software for multi-track recording and editing.

The pseudoword stimuli consisted of six formants – /e/, /oʊ/, /ɪ/, /i:/, /u:/, and /ɑ:/ – within four pseudowords – H—T; F—T; D—T; K—T – for a total of 24 stimuli. As planned the fourth word in each line consisted of the K—T construction, and was not used in the analyses due to the natural tendency for falling intonation during recitation. The remaining three constructions – H—T; F—T; and D—T – were randomised between each line in two iterations of the stimulus presentation. That is, two copies of the stimulus paper were generated. The stimulus paper was standard letter sized white paper in 28-point Constantia font, chosen for the clear distinction between letters.

### 2.3 Data Recording and Processing

[2] At the time of writing, the DAL is in development for digital assessments and tabulations.

Audio recordings were made with a Samson C01U USB Studio Condenser Microphone (Samson Technologies, Hauppage, NY) with an operational frequency response of 20-18000 Hz and a 19mm diaphragm with a maximum Sound Pressure Level of 136 dB, capable of digital recordings of up to 16-bit, 48-kHz output. The microphone was operated with a Lenovo IdeaPad Y580 laptop with Microsoft Window 10 operating system through the USB 3.0 port and facilitated by the Audacity (©2016 Audacity Team) open-source digital audio workstation for recording. Recordings were made at the maximum 16-bit; 48-kHz allowance to maximize data collection and while the laptop remained plugged in as to prevent unwanted power drain. Individual recordings were exported as high quality mp3 files for further data processing.

The mp3 files of the subjects' audio recordings were then imported into PRAAT Phonological Analysis software (Boersma & Weenink, version 5.3.86, September 2014). This open source phonetics software was used to quantify the physical descriptors of each of the target stimulus, which were Minimum Pitch, Maximum Pitch, Mean Pitch, Minimum Intensity, Maximum Intensity, Mean Intensity, and the first four Formants (F1-F4). These values were gathered for each stimulus and entered into an SPSS database for further processing and analyses. SPSS was used to calculate additional variables such as Pitch Range, Intensity Range, and the distances between all formant combinations, *i.e.* F2-F1, F3-F1, F4-F1, F3-F2, F4-F2, and F4-F3.

The *pitch* of a human voice is quantifiable as the fundamental voice frequency, or *F0*. It is the most sexually dimorphic quantum of the human voice as it is a direct measure of the vibrational frequency of the vocal chords (Puts, Gaulin, & Verdolini, 2006). While *pitch* reflects the frequency of the sound produced, *intensity* is a quantum of the sound pressure level (volume) of the sound. Clinical research has shown not only a strong correlation between *F0* and the intensity of speech, but these can be used as diagnostic criteria (Hirano, Vennard, & Ohala, 1970; Komiyama, Watanabe, & Ryu, 1984). Voice intensity has also been linked with more complex forms of human verbal behaviour, such as deception (Ekman, Friesen, & Scherer, 1976). Finally, the four formants (*F1-F4*) link intimately with the fundamental frequency of the individual, but occur as the result of associated speech structures in addition to the vocal folds. Research has demonstrated that the first two formants are typically sufficient to distinguish any two vowels, in particular the ratio of F1:F2, known as *vowel space* (Neel, 2008).

### 2.4 Statistical Analyses

Statistical analyses were carried out using IBM SPSS Statistics v23 for Windows operating systems, and all errors displayed in the figures to follow are standard errors of the mean.

## 3. Results and Discussion

The results will be organized according to the dependent variable focussed on in each analysis.

### 3.1 Fundamental Frequency

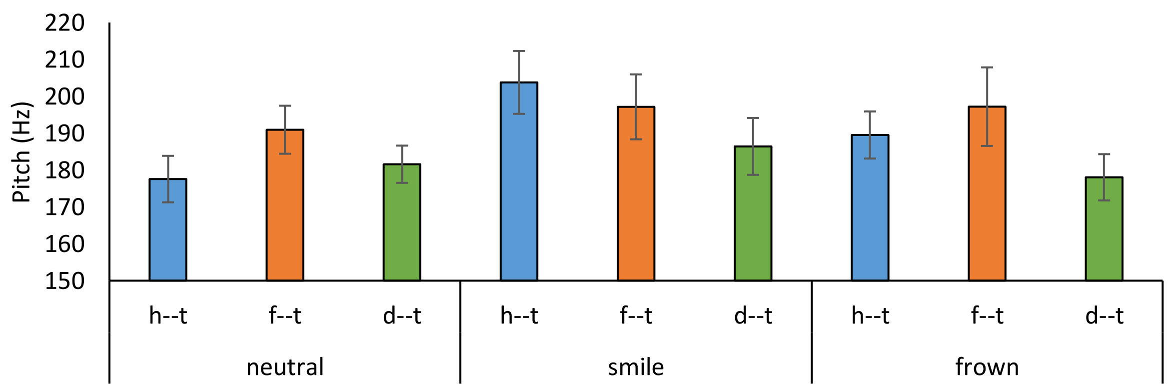


**Figure 1.** The con*word interaction effect.

The results of the generalized linear model with pitch as the dependent variable and including gender (gen), facial expression (con), pseudoword type (word), and target vowel (vow) variables indicated significant main effects for gen (F=34.072, partial $\eta^2$=0.709), con (F=7.208, partial $\eta^2$=0.34), vow (F=23.885, partial $\eta^2$=0.63), word (F=3.375, partial $\eta^2$=0.21), and a two-way interaction between con*word (F=2.91, partial $\eta^2$=0.172). As expected, males displayed lower mean fundamental frequency compared to females (152.743 Hz vs. 225.493 Hz; not shown). Figure 1 displays the con*word interaction, reflecting a potential pattern that smiling increases the pitch of vocalization with a more word-conditional increase when frowning. Of interest, only the H—T and D—T pseudowords appear to show significant pitch shifts; those for the F—T pseudoword were not significantly different with respect to facial expression.

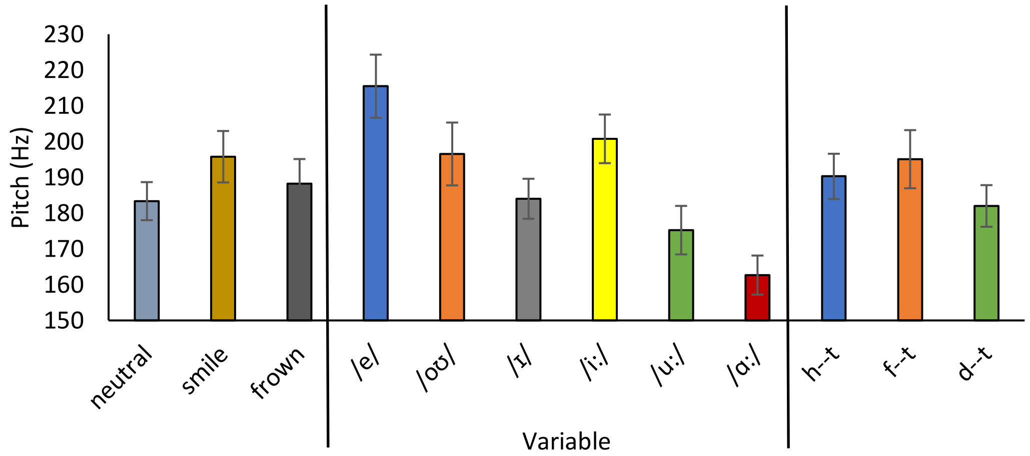


**Figure 2.** the collected main effects on Fundamental Frequency separated by vertical black bars**.**

Figure 2 displays the three main effects revealed for con, vow, and word, showing a possible trend that both facial expressions increased the pitch of vocalization, but contrary to Figure 1 it suggests that the first two pseudowords H—T and F—T were significantly higher than the D—T construction. Of most interest is the different effects on pitch that the individual formants had on pitch, possibly reflecting a point of articulation effect; in general, more forward formants had a higher pitch and those articulated near the back of the oropharynx had lower pitches.

**3.2 Intensity**

The results of the generalized linear model with overall intensity of vocalization as the dependent variable and including gender (gen), facial expression (con), pseudoword type (word), and target vowel (vow) indicated no significant main effects nor interactions with respect to the mean intensity produced in the vocalizations.

*Formant One*

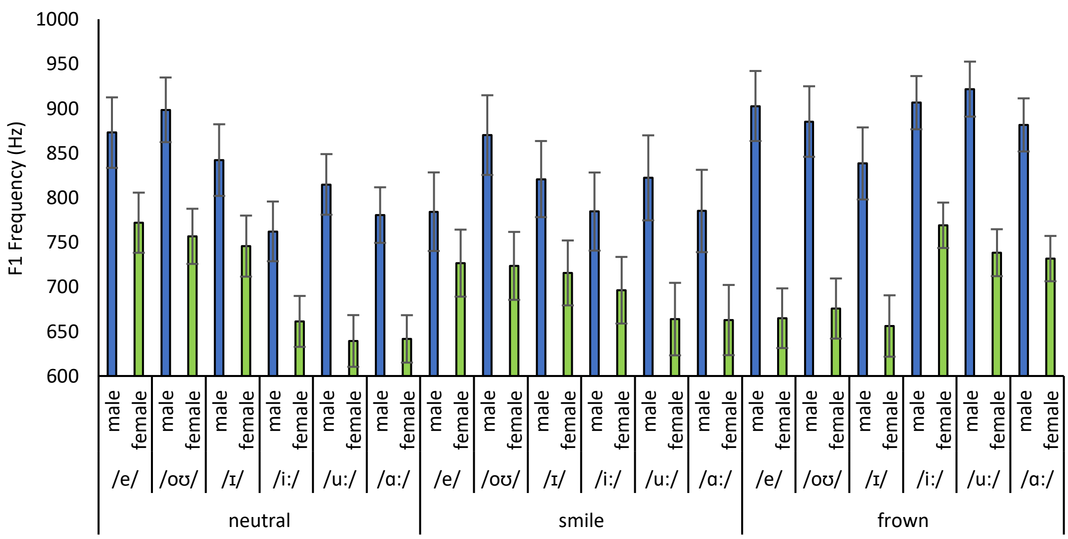


**Figure 3.** the three-way interaction between gen*con*vow

P ISSN: 2621-0843
E ISSN: 2621-0835



The results of the generalized linear model with the First Formant as the dependent variable including gender (gen), facial expression (con), pseudoword type (word), and target vowel (vow) indicated significant main effects on formant one with respect to gen ($F=9.431$, partial $\eta^2=0.357$), con ($F=5.880$, partial $\eta^2=0.251$), vow ($F=8.903$, partial $\eta^2=0.344$), and word ($F=3.651$, partial $\eta^2=0.177$), as well as determining a two-way interaction between con*vow ($F=27.220$, partial $\eta^2=0.616$) and a three-way interaction between gen*con*vow ($F=3.691$, partial $\eta^2=0.178$). Perhaps most interestingly, females showed the lower mean Formant 1 frequency as compared to males (702.184 Hz vs. 842.924 Hz), an effect which pervades the three-way interaction, as shown in Figure 3 which breaks down the analysis into the three-way interaction. The effect of the vowel itself presumably due to point of articulation is still somewhat present, though it appears masked by smiling and somewhat inversed for the frown. The male-female difference is more apparent in particular word-vowel combinations than others, though females uniformly have lower Formant 1 Frequencies.

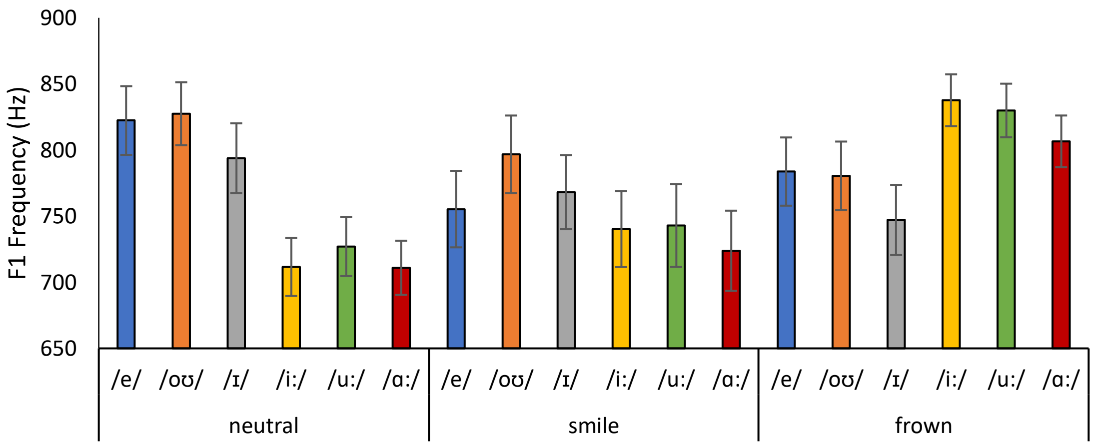


**Figure 4.** the two-way interaction between con*vow.

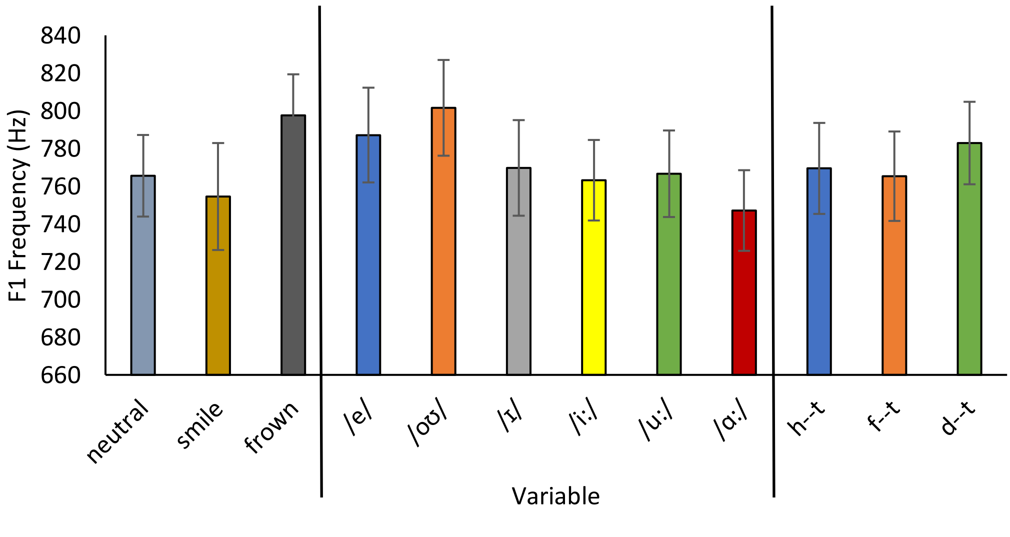


**Figure 5.** the collected main effects on Formant 1 separated by vertical black bars

The two-way interaction appears in Figure 4 which again illustrates a point of articulation effect for Formant 1, but one that is dependent on the facial expression. The forward-back differences are most apparent for the neutral facial expression, while smiling seems to remove much of the significant differences between vowels within Formant 1. However, an inversion of frequency peaks occurs when participants were frowning. Figure 5 demonstrates the power of this inversion, where in general frowning induced a shift to higher frequencies in Formant 1, an effect also seen in the pseudowords where H—T and F—T did not differ from one another yet D—T in general increased Formant 1 frequencies. The point of articulation effect is less visible within this analysis but is still suggested by direct comparison of /oʊ/ with /ɑː/.

*Formant Two*

The results of the generalized linear model with Formant Two as the dependent variable and including gender (gen), facial expression (con), pseudoword type (word), and target vowel (vow) indicated significant main effects for gen (F=10.800, partial $\eta^2$=0.388), con (F=53.968, partial $\eta^2$=0.760), vow (F=47.817, partial $\eta^2$=0.738), and word (F=4.290, partial $\eta^2$=0.202). The generalized linear model also indicated two significant two-way interactions vow*gen (F=2.760, partial $\eta^2$=0.140) and con*vow (F=40.722, partial $\eta^2$=0.705), as well as a three-way interaction between con*vow*gen (F=5.917, partial $\eta^2$=0.258). Females once more show the lower mean Formant Two frequency as compared with males (1887.966 Hz vs. 2009.961 Hz), an effect pervading the three-way interaction although not as powerfully as within Formant 1, although particularly prominent in the neutral and smiling condition and vowels articulated near the back of the oropharynx and in the frowning condition for forward-articulated vowels (Figure 6).

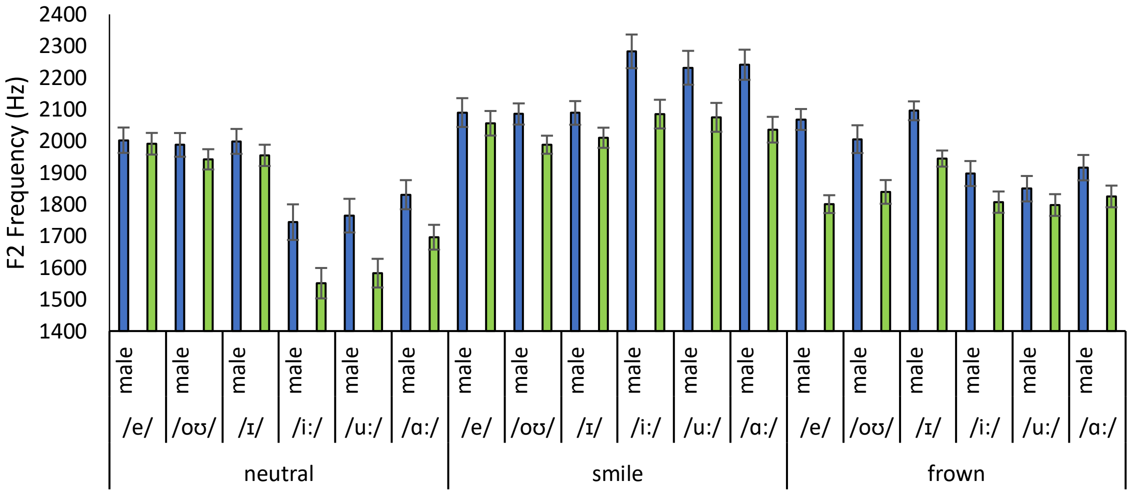


**Figure 6.** the three-way interaction between gen*con*vow

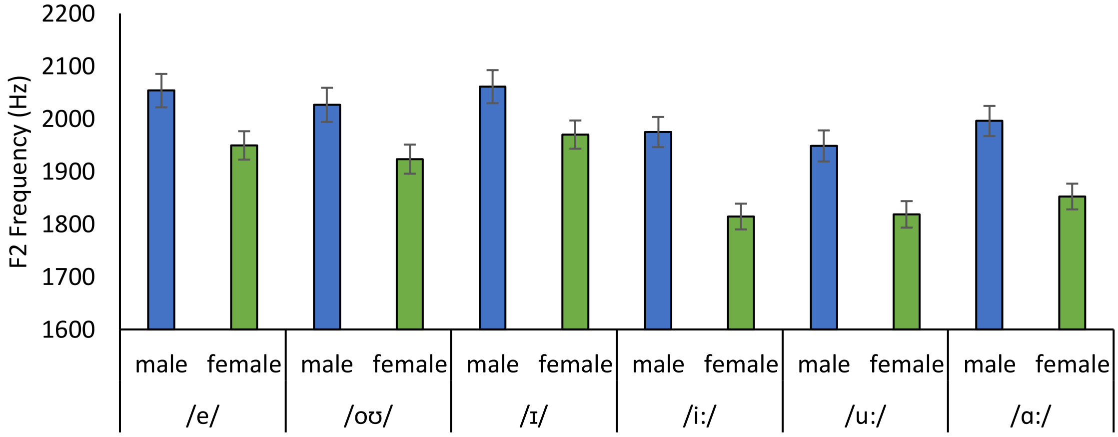


**Figure 7.** the two-way interaction between vow*gen

The two-way interaction between vow*gen follows in Figure 7 which further illustrates that males and not females vocalized with higher Formant 2 Frequencies for each of the target vowels. Figure 8 shows the con*vow interaction effect, where it provides further evidence of potential point-of-articulation effects, as in the neutral condition there is an incredible drop in Formant 2 frequency for those vowels produced at posterior oropharynx, while the inverse is true while smiling. No consistent pattern is discernable while frowning. Figure 9 displays the collated main effects, where an impressive effect of an upward shift in Formant 2 frequency occurs during the smiling condition, the anterior vs. posterior point-of-articulation effect is again quite apparent, and only a subtle effect of pseudoword is visible irrespective of other variable interactions.

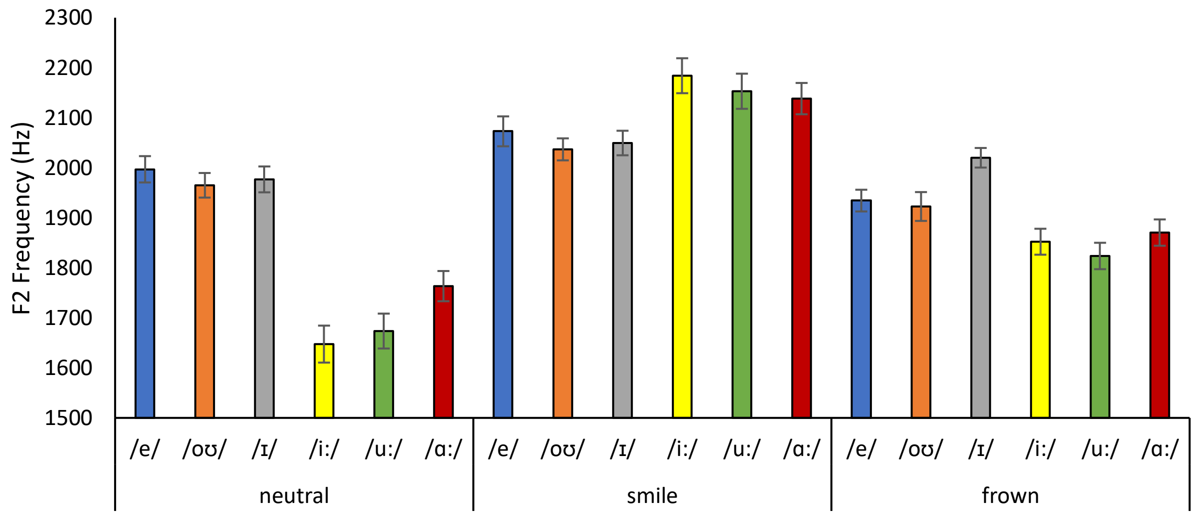


**Figure 8**. the two-way interaction between con*vow

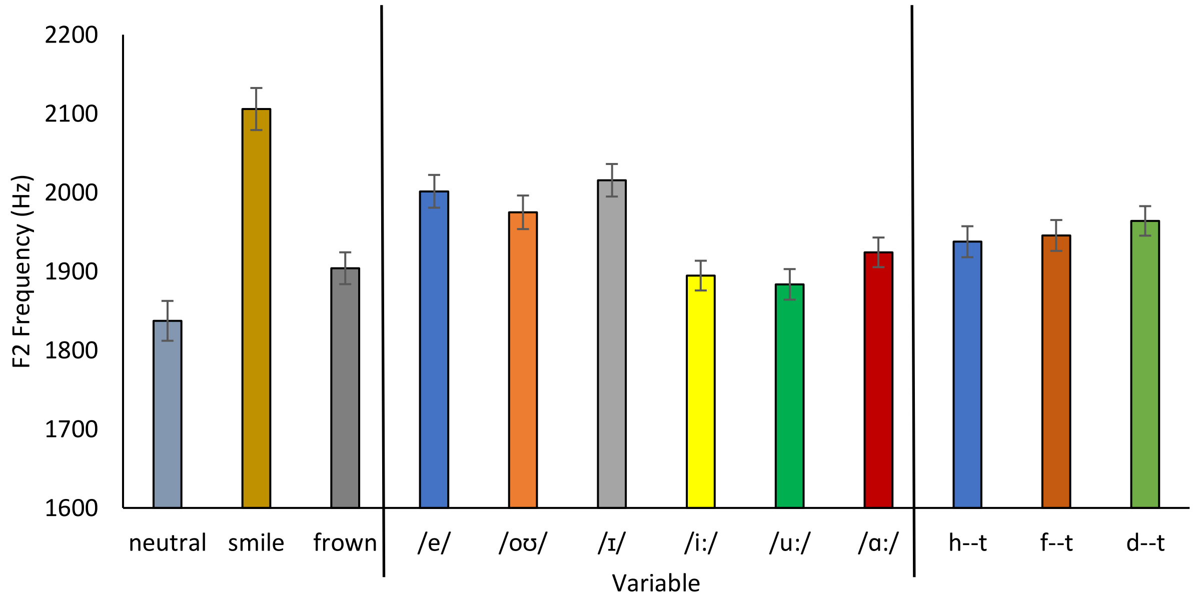


**Figure 9.-** the collected main effects on Formant 2 separated by vertical black bars**.**

## 4. Conclusion

The central thrust of the present work was to investigate the interaction between emotion as expressed in facial expressions and the physical sound properties of vocalizations; that is, the anecdotal effect of "sounding happy or sad." While there are of course more powerful and obvious sources of variance that contribute to the ultimate sound properties of a vocalization in any given circumstance, the present paper sought to explore the contribution of the facial musculature. The results resoundingly point to a strong interaction between the musculature of the face and the sound properties of vocalizations, albeit no consistent pattern has emerged. Of the numerous effects, those expected were the general increase in pitch during the "smiling" facial expression corresponding to a happy mental state, where previous research has suggested that angry/happy emotional states leads to a widening of Pitch Range and an increase in Intensity, and the opposite for sad/bored states (Murray & Arnott, 1993). The present results would thus seem to corroborate the findings of an increase in Pitch Range associated with happy states (*smiling*); however, the results are somewhat conflicting as "frowning," a facial expression commonly linked with a "sad" mental state, also lead to a general increase in pitch, with the exception of the d—t pseudoword. While somewhat contentious, these results corroborate previous studies that have correlated affective states with Pitch Range (Cahn, 1990; Roy & Pentland, 1996; Scherer, Koivumaki, & Rosenthal, 1972).

Similarly, the expected male vs. female dichotomy occurred in Fundamental Frequency, with males having generally lower F0's than females, which may be in line with literature suggesting the Fundamental Frequency as being a potential contributor to conveying affective states in speech (Allen, Burton, Olman, & Oxenham, 2016; Cahn, 1990). Further studies are necessary to determine the true extent that specific sounds influence the emotional state of both speaker and listener.

The vowel/phoneme specific differences are of note but when viewed in the context of speech are to be expected – the physical differences in the sounds of phonemes is what enables the classification into different phonemes, lest we have languages of allophones (Wells, 1945). Accounting for these natural variations, the present study found significant interactions between facial expressions, words, and the vowels vocalized within Formants One and Two, including significant gender interactions, perhaps corroborating previous findings and those reported above (Murray & Arnott, 1993). However, the most unexpected effect within both formants was that males displayed higher mean frequencies in all conditions compared to females, an effect that was exaggerated dependant upon facial expression and vowel. It is possible that this may be due to interactions between point-of-articulation of the vowels and the facial expressions/musculature as well as general muscle exertion; holding these particular facial expressions alters were vowels are produced, albeit subtly, but enough that it may reflect the emotional state of the individual producing them. Although we expected for formants to change with respect to vowel, it was not expected that male formant frequencies would be universally higher than females; despite this surprise, the phenomenon has been detailed in the literature previously (Bennett, 1981; Busby & Plant, 1995; Huber, Stathopoulos, Curione, Ash, & Johnson, 1999). Thus, our findings corroborate these earlier studies and also emphasize the interplay with emotional or affectual context, although once more we emphasize further and more rigorous studies are necessary to elucidate the interplay of sound and emotional valence of speaker/listener.

Despite this with some certainty it is clear that the gender variable is a very large source of variance, contributing to the dispersion of the Fundamental and Formant Frequencies, although it is somewhat intriguing no intensity effects were present. Although we did not pursue discriminant functions here, a number of previous works have demonstrated strong classification accuracies ranging from 50-90% using identical or similar variables as those measured in the present study (Childers, Wu, Bae, & Hicks, 1988; Kotti & Kotropoulos, 2008; Sedaaghi, 2009; Ververidis & Kotropoulos, 2004). Taken together, it would seem to support the overt effects of sexual dimorphism on the physical structure of the vocalization apparatus primarily, but coupled with the interaction effects may also suggest a dichotomous ability in the production or even mimicry of emotional states.

Whissell's work demonstrates that at least some aspect of emotional affect is carried within words themselves, and more specifically that this inherent affective domain is a result of the composite phonemes of each word (Whissell, 2000, 2009). The present study would seem to corroborate this finding, as the word main effect occurred for Fundamental Frequency, Formant One, and Formant Two, which together comprise the most important constituents of speech, in addition to the interaction effects. Although it is difficult to surmise a pattern in a specific manner – *i.e.*, that h—t is a sadder sound compared to f—t or that one always results in a lowering of frequency – suffice it to say that the present work demonstrates the complex interaction between sound and emotional state as likely reflecting a deeply structured system, much as postulated and refined by Chomsky (N. Chomsky & Lightfoot, 2009). While the present study may indirectly support Whissell's work, one requires a more rigorous design with the explicit objective of demonstrating the phonoemotional profile of phonemes to truly elucidate the interaction effects discussed above. The present study does show that some sounds are more emotionally salient than others, although whether the affect arises from the sound or the intention of the speaker remains in question.

When engaging in communicative behaviour the goal is to alter the mentation of the other individual, whether through the imparting of knowledge or sharing of experiences (Frith & Frith, 2006). However, the successful impartation of knowledge requires a listener to be able to discern the speaker's intent (Sperber, Wilson, 何自然, & 冉永平, 1986). In the context of this experiment, there was no real intent from the speakers to communicate an underlying affective state, aside from complying with the wishes of the experimenter. Although volunteers are attempting to mimic the mental states of neutral, smiling, or frowning individuals, ultimately their intent does not change during the task. Significant differences in brain activation within the dorsal medial frontal cortex and ventral prefrontal cortex indicate to what extent *intent* plays a role in metacognition for communication (Frith & Frith, 2006; Mitchell, Macrae, & Banaji, 2006). In addition, speech aspects such as prosody and melody may carry this communicative intent, as demonstrated by research on mothers' speech to infants (Fernald, 1989). Taken together, these studies suggest that the interpretation of the present manuscript be with respect to the mechanistic effects of the facial muscles, through holding facial expressions, on the vocal tract and sounds it produces, rather than underlying affective states targeted by the expressions.

When considered from the context of **COMMUNICATIVE INTENT** the present work establishes an interesting prescience in the philosophy of language. The pseudowords used in the study are, by definition, meaningless, thus arguing more in favor of the mechanistic explanation noted above. Communicative intent is impossible if the medium used to communicate cannot carry meaning. Yet despite this apparent 'meaninglessness' of pseudowords, through the context of the experiment and the participation of volunteers, the words have acquired *some* meaning. HEHT does not have a definition, but the experimenter and participants now have a **REFERENT** for this non-word, an associated explanation to go along with it, memories, and contexts where it does (or did for a time) "make sense." The non-word is thus usable in numerous language-games by the experimenter and participants, as analogously as Wittgenstein's 'beetle' (Wittgenstein, 2010). If this holds true then are words any different, or are they excluded from the principle that meaning is derivable at whim through use and context? Words it seems are placeholders for future semantic information storage and retrieval; a number of anthropological works on the development and evolution of language would seem to suggest so (Blasi et al., 2019; Hodgson, 2019; Ramiro, Srinivasan, Malt, & Xu, 2018; Taylor, Davis, & Rastle, 2019).

Overall, the present study demonstrates a significant and complex interaction between the musculature of the vocal apparatus and the emotional-affective system, much in the way many cultures anecdotally state that one "sounds happy." However, by far gender differences and interactions dominated much of the results; despite this, it was determined conclusively that different facial expressions do in fact alter the phonological properties of speech, as demonstrated by the significant differences in Fundamental, Formant One, and Formant Two frequencies in line with previous research. There interactions between words, vowels, and facial expressions also lend credence to Whissell's phonoemotional theories of language, synthesizing with ideas put forth as well by Ekman and Chomsky, reflecting a deep, perhaps even convoluted, structure of language. Further research is required to explore to what extent a 'semantic' or 'meaning' dimension for words may exist, as evidenced by other recent studies, and whether communicative intent can occur with 'meaningless' words (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016a; Kocagoncu, Clarke, Devereux, & Tyler, 2016). Evidence exists supporting the independence of speech sounds from their composite words, demonstrating heavily influence from the environment, thus suggesting physical referents for these sounds (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Maddieson & Coupé, 2015; Regier, Carstensen, & Kemp, 2016). It appears as though both sounds and their composites words can be both meaningful and meaningless, dependent heavily on the context and situation in which they occur in; and thus Socrates' middle ground on the origin of words combines brilliantly with Aristotle's compromise on the subject, and rings as true today as it did 2500 years ago (Ackrill, 1975; Aristotle, 2009; Modrak, 2001; Plato, 2015).

**Acknowledgments**

We would like to thank the Neuroscience Research Group for their support and assistance with organizing the study and for internal review of the paper, specifically Dr. Blake Dotta and the late Dr. Michael Persinger.

**References**

Ackrill, J. L. (1975). *Categories and De Interpretatione*: Clarendon Press.

Allen, E. J., Burton, P. C., Olman, C. A., & Oxenham, A. J. (2016). Representations of Pitch and Timbre Variation in Human Auditory Cortex. *Journal of Neuroscience*, 2336-2316.

Aristotle. (2009). *The Basic Works of Aristotle* (R. McKeon, Trans.): Modern Library.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America, 69*(1), 231-238.

Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science, 363*(6432), eaav3218.

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*, 201605782.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America, 97*(4), 2603-2606.

Cahn, J. E. (1990). The generation of a ect in synthesized speech. *Journal of the American Voice I/O Society, 8*, 1-19.

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. *Darwin and facial expression: A century of research in review*, 11-89.

Childers, D., Wu, K., Bae, K., & Hicks, D. (1988). *Automatic recognition of gender by voice.* Paper presented at the Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language, 35*(1), 26-58.

Chomsky, N., & Lightfoot, D. W. (2009). *Syntactic Structures*: De Gruyter.

De Saussure, F. (2011). *Course in general linguistics*: Columbia University Press.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), 124.

Ekman, P., Friesen, W. V., & Scherer, K. R. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica, 16*(1), 23-28.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology, 30*(1), 527-554.

Feld, S. L. (1981). The focused organization of social ties. *American journal of sociology, 86*(5), 1015-1035.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child development*, 1497-1510.

Fine, C. (2008). *The Britannica Guide to the Brain: A Guided Tour of the Brain - Mind, Memory, and Intelligence*: Robinson.

Frege, G. (1948). Sense and reference. *The philosophical review, 57*(3), 209-230.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience, 37*(18), 4848-4858.

Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch and intensity of voice. *Folia Phoniatrica et Logopaedica, 22*(1), 1-20.

Hodgson, D. (2019). The origin, significance, and development of the earliest geometric patterns in the archaeological record. *Journal of Archaeological Science: Reports, 24*, 588-592.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America, 106*(3), 1532-1542.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016a). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*(7600), 453-458.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016b). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*(7600), 453.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron, 76*(6), 1210-1224.

Jaynes, J. (2000). *The Origin of Consciousness in the Breakdown of the Bicameral Mind*: HMH Books.

Kocagoncu, E., Clarke, A., Devereux, B., & Tyler, L. K. (2016). Decoding the cortical dynamics of sound-meaning mapping. *Journal of Neuroscience*, 2858-2816.

Komiyama, S., Watanabe, H., & Ryu, S. (1984). Phonographic relationship between pitch and intensity of the human voice. *Folia Phoniatrica et Logopaedica, 36*(1), 1-7.

Kotti, M., & Kotropoulos, C. (2008). *Gender classification in two emotional speech databases.* Paper presented at the Pattern Recognition, 2008. ICPR 2008. 19th International Conference on.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review, 94*(4), 412.

Maddieson, I., & Coupé, C. (2015). *Human language diversity and the acoustic adaptation hypthesis.* Paper presented at the Proceedings of Meetings on Acoustics 170ASA.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*(4), 655-663.

Modrak, D. K. W. (2001). *Aristotle's Theory of Language and Meaning*: Cambridge University Press.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America, 93*(2), 1097-1108.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research, 51*(3), 574-585.

Pestilli, F. (2018). Human white matter and knowledge representation. *PLoS biology, 16*(4), e2005758.

Plato. (2015). *Cratylus* (T. Taylor, Trans.): CreateSpace Independent Publishing Platform.

Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*(4), 283-296.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences, 115*(10), 2323-2328.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: words for snow revisited. *PloS one, 11*(4), e0151138.

Roy, D., & Pentland, A. (1996). *Automatic spoken affect classification and analysis.* Paper presented at the Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*(6), 2408-2412.

Scherer, K. R. (1985). Vocal affect signaling: A comparative approach. *Advances in the study of behavior, 15*, 189-244.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research, 1*(3), 269-285.

Sedaaghi, M. (2009). A comparative study of gender and age classification in speech signals. *Iranian Journal of Electrical and Electronic Engineering, 5*(1), 1-12.

Simonyan, K., Ackermann, H., Chang, E. F., & Greenlee, J. D. (2016). New Developments in Understanding the Complexity of Human Speech Production. *Journal of Neuroscience, 36*(45), 11440-11448.

Skinner, B. F. (2014). *Verbal behavior*: BF Skinner Foundation.

Sperber, D., Wilson, D., 何自然, & 冉永平. (1986). *Relevance: Communication and cognition* (Vol. 142): Citeseer.

Taylor, J., Davis, M. H., & Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. *Proceedings of the National Academy of Sciences*, 201818575.

Ververidis, D., & Kotropoulos, C. (2004). *Automatic speech classification to five emotional states based on gender information.* Paper presented at the Signal Processing Conference, 2004 12th European.

Wells, R. S. (1945). The pitch phonemes of English. *Language*, 27-39.

Whissell, C. (1989). The dictionary of affect in language. emotion: theory, research and experience 4. The Measurement of Emotions. In: Academic Press, New York.

Whissell, C. (2000). Phonoemotional profiling: a description of the emotional flavour of English texts on the basis of the phonemes employed in them. *Perceptual and Motor Skills, 91*(2), 617-648.

Whissell, C. (2001). Sound and emotion in given names. *Names, 49*(2), 97-120.

Whissell, C. (2006). The flow of emotion through Beowulf. *Psychological reports, 99*(3), 835-850.

Whissell, C. (2009). Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological reports, 105*(2), 509-521.

Wittgenstein, L. (2010). *Philosophical investigations*: John Wiley & Sons.