THE VOICE OF LOVE

Trisha Belanger, Caroline Menezes, Claire Barboa, Mofida Helo, Kimia Shirazifard

University of Toledo, United States
tbelanger@rockets.utoledo.edu, Caroline.Menezes@utoledo.edu, Claire.Barbao@rockets.utoledo.edu, Mofida.Helo@rockets.utoledo.edu, Kimia.Shirazifard@rockets.utoledo.edu

ABSTRACT

The present study is a phonetic analysis of the emotions of love. We focus on the voicing differentiation of romantic love, parental-child love, friendship (love for a friend), and patriotism. Open quotient, spectral tilt, formant frequencies and amplitudes were extracted from a target word spoken by male and female college students as they provided the voice over for selected video clips. Results reveal gender differences for open quotient and fundamental frequency. However, none of the voice source parameters were clearly able to differentiate the different forms of love.

Keywords: Spectral tilt, open quotient, love, emotion, voice quality

1. INTRODUCTION

We are well aware of how vocal expressions of emotions influence our social interactions [14] but few studies have shown how we adapt our voice to our relationship to the speaker. Human expressions of love, as we know, appear to be strongly influenced by social relationships. Love between a child and parent differs from patriotism or love between two people that are romantically involved [13]. Love is universal, multifaceted and vital to the existence of humans [17].

Speech is acoustically rich because it conveys both the linguistic message as well as considerable personal information about the talkers [1]. A recent study [6] explored the role of fundamental frequency and voice quality in the identification of attitudes and emotion of French speakers. In this experiment 22 subjects were asked to imagine themselves in the specified emotional state or attitude and were to read three different texts containing a neutral sentence within the middle. It was theorized that each neutral utterance would hold emotionally charged variations within the voice that could help identify the emotional state or attitude of the speaker. Results indicated that attitudes could be perceived through prosodic contour, while emotions were associated with both prosodic contour and voice quality.

Phonetic studies have reported variations in prosodic contours and voice quality for different emotions such as anger, fear, happiness and sadness [2, 34, 8, 9, 10, 11, 12, & 15]. Even synthesized voices that were systematically varied for voice quality have been shown to convey differences in affect [5].

But, is it possible to determine the relationship of the speakers based on their voice characteristics? Can prosodic and voice characteristics distinguish social relationships between the speaker and listener, in particular the different forms of love.

Methodologically it is difficult to collect systematically controlled emotional speech within the laboratory. Collecting accurate data on intimate emotions is even more problematic. To be able to distinguish the distinctions between the different forms of love we have to be able to tease out small variations in speaker's acoustic signal. According to [5], voice quality analysis is better able to parse these mild states when compared to broad prosodic features like rhythm and intonation analyses.

Synthetic speech or portrayals of emotion by professional actors have provided most research on vocal expression of emotion. One such experiment that investigated the vocal parameters to distinguish the emotions of 'neutral', 'sadness', 'joy', and 'tenderness' was conducted on Finnish actors [17]. Nine professional actors produced 10 renditions of each emotion as they read a passage from a novel. Their results indicated that linguistic context is necessary for the accurate differentiation of vocal emotion. According to these researchers, emotions are usually replicated in order to produce the emotion and laboratory recordings capture only differences in the replication of the emotion. Acted speech samples are similar to natural speech, but hold the idea that actors' imitations may be influenced by stereotypes of emotional expression and may showcase intensified qualities of speech affecting the perception of listeners.

What sets this current study apart is that the emotions measured were not imitated by the professional actors, but they were recorded from ordinary young college students. Our experimental paradigm was designed to have our speakers feel the emotion by having them observe an emotional video and by asking subjects to dig into prior feelings, related to the emotion in focus. We also maintained the target word and vowel but varied the utterances based on the target emotion.

This paper is a preliminary study to test the phonetic correlates of four different forms of love, namely, romantic (RL), parental (PCL), friendship (FSH) and patriotism (PTR). It tests a novel paradigm for recording speech that expresses natural emotion within the laboratory. Voice quality measures were analyzed and compared for the different types of love.

2. METHODS

2.1. Participants

A total of 36 subjects, 13 males and 23 females, from American college students were recorded. All subjects were students and voluntary participants in the study. Subjects were not told the exact purpose of the study but they were all aware that they would have to provide the voice for famous Disney characters.

2.2. Procedure

In this study, emotional speech was recorded in a paradigm where subjects were to assume the role of a "voice actor." The experiment was designed and intended to stimulate and surface authentic emotions in subjects within the laboratory. To this purpose, subjects were instructed to watch short clips from popular animated movies (The *Lion King, Aladdin* and *Shrek)*. The movie clips were approximately a minute long and conveyed content indicative to one of the four different kinds of love (*romantic love, friendship, patriotism, and parental-child love*), anger, and fear. The scenes were carefully chosen and spliced using iMovie software. Subjects watched these clips in a quiet, dark room on a computer screen. After watching a given clip, they were asked to revisit incidents in their life relevant to how the clip made them feel. Subjects were not told what emotion to produce for any of the video.

Subjects were provided an index card with a written scenario and a composed script of what they were to say. A shortened version of the clip was then shown if needed, followed by a still shot at which time the subject was expected to record their utterance. Most of the subjects chose to read over the script several times. They were allowed to record as many times as needed until they felt they had done their best. For each clip there was one sentence that the subject had to speak using his/her best rendition of that emotion. Every sentence began with the words "I need you" followed by a string of 6-7 words that were a close approximation of the movie content. The phrase "I need you" was considered the target phrase and "need" was the target word containing the target vowel /i/. All recordings were made using the Marantz Professional Portable Solid State Recorder Device. All recordings were recorded as .mp3 files and then converted to .wav files using the phonetics software program PRAAT developed by Paul Boersma. Only recordings of the four different loves were analyzed in this paper.

2.2 Acoustic Analysis

The word "need," with the /i/ phoneme was isolated using the textgrid function in the phonetic software Praat. VoiceSauce (an open source phonetic analysis program developed by UCLA Department of Linguistics) was then used to automatically extract the relevant acoustic parameters.

The acoustic parameters analyzed included the amplitudes of the first harmonic (H1) and second harmonic (H2) along with the harmonic of the third formant (A3). VoiceSauce also automatically calculated the open quotient and spectral tilt values. Open quotient is the difference between the amplitude of the first and second harmonics (H1-H2). Spectral tilt was calculated as the difference between the amplitude of the first harmonic and the harmonic associated with the third formant (H1-A3). Since the vowel used in this study was only the front high vowel /i/ no normalization was done on the data to account for the effect of fundamental frequency on the first formant.

According to [7] open quotient relates to distance of the left and right vocal fold at the time of the open phase, which would directly relate to the breathiness of the utterance. Spectral tilt on the other hand indicates the rate at which the vocal folds come together in the closing phase. The rate at which the vocal folds close is dependent on the velocity of the air particles flowing through the opening. At higher velocities the vocal folds close at a quicker rate.

3. RESULTS

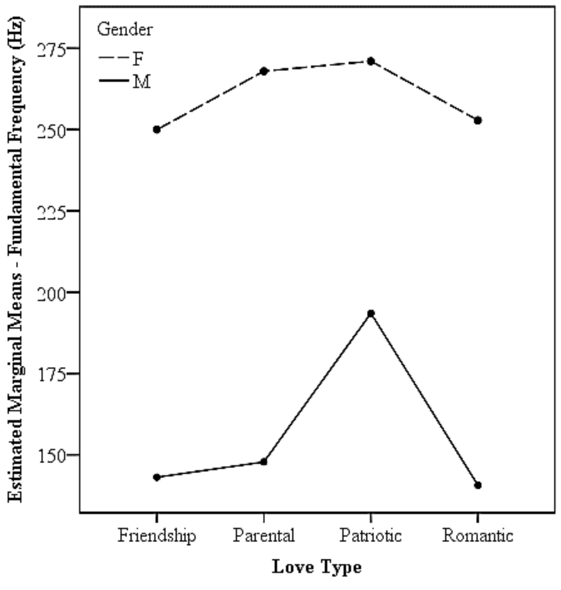
Of a total of 148 utterances, eleven utterances were omitted because they returned 0 values in the automatic extraction process. (See Table 1 for the number of utterances that were analysed.)

Table 1 gives the number of utterances, the mean, and standard error for the following variables: amplitude of the H1, H2, A3, fundamental frequency, open quotient, and spectral tilt measures for all speakers separated by gender and love type.

3.1. Fundamental Frequency

Fig. 1 plots the estimated marginal means for fundamental frequency separated by gender and love type. Estimated marginal means were used in this study because of the differences in sample size. As is expected, females have significantly higher fundamental frequency than males ($F(1,128)= 234.2$, $p<.001$). Both genders tended to have lower fundamental frequency for FSH and RL and higher fundamental frequency for PCL and PTR. Love type was also significantly different ($F(3,128) = 6.0$, $p=.001$) however the partial Eta squared values was very low ($\eta^2=.123$) indicating a very small effect. The Tukey post hoc test revealed that F0 for PTR was significantly higher than FSH and RL respectively ($p<.01$).

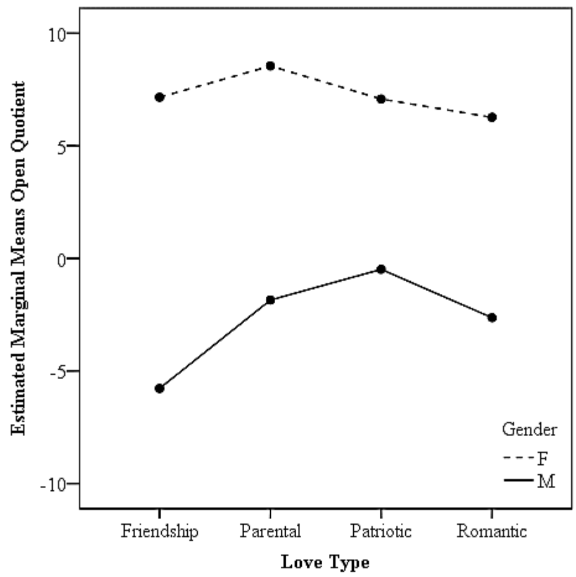
Figure 1: Estimated marginal means for fundamental frequency separated by gender and love type.



3.2. Open quotient

Fig. 2 plots the estimated marginal means for open quotient values (H1-H2) for the different love types separated by gender. High positive values would indicate that the amplitude of the first harmonic was higher than the amplitude of the second harmonic as is seen in breathy voices where the vocal folds are further apart from each other during the open phase. Again we see that females have higher open quotient values when compared to males. Univariate ANOVA analysis revealed that this difference is significant at $F(1,128) = 101$, $p<.001$. This would mean that the vocal folds are significantly more open in females than males at the open phase. In this data, we observe that women use a more breathy voice when speaking with their children but men unexplainably use this voice when expressing their love for their country! There were no significant differences however, between the love types and no significant interactions between gender and love type.

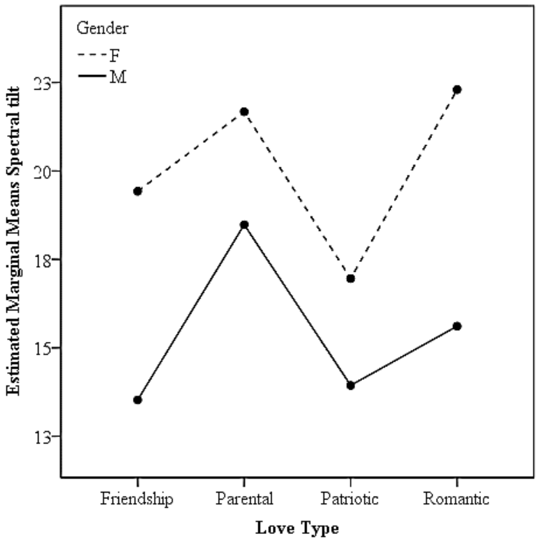
Figure 2: Estimated marginal means for open quotient separated by speaker and love type.



3.3. Spectral tilt

Fig. 3 plots the estimated marginal means for spectral

Figure 3: Estimated marginal means for spectral tilt separated by speaker and love type.



tilt (H1-A3) for the different types of love separated by gender. Higher values would indicate a great spectral tilt resulting from lower air velocities during the closing phase. Univariate ANOVA analysis again revealed a significant though small gender effect but no effect was observed for love type ($F(1,128) = 9.50$, $p=.003$; $\eta^2=.069$). No significant relationship between gender and love type was also observed.

4 CONCLUSION

This study attempted to distinguish the voice quality of love in a systematic and controlled study within the laboratory. As highlighted by Gobl & Chasaide [5] the study of voice has often been neglected due to methodological and conceptualization difficulties. Production studies have been particularly hampered by the difficulty in recording natural spontaneous emotional utterances. Researchers either study emotion portrayals from actors or use ethically acceptable stimulation techniques, resulting in fairly weak affect states in the subjects.

What sets this study apart is that we attempted to test the general populations' perception of how these emotions are produced while trying to avoid bad replications of the true emotion. To enhance their acting abilities we dug into the subjects' prior feelings by asking them to recall previous experiences. However, we are not able to know if subjects truly delved into their past. We also controlled the target utterance carefully while still keeping the entire linguistic utterance emotionally relevant.

Our findings were not unique in reporting gender differences in voicing, but this study also showed that the open phase of a glottal cycle was more relevant than the closing phase for differentiating small differences within an emotion. Larger open quotients were found in female speakers than male speakers indicating that females use a more breathy voice when speaking in love. While there were gender variability for the four types of love these differences were not found to be significant.

Fundamental frequency analysis also revealed that males and females use lower pitch when speaking with their friend or loved one, and higher pitch when speaking with children or of their country.

The results in this study indicate that the variation between the four love types study were negligible. It is possible that more data might be needed to parse out these small differences but it is also possible that finer distinctions in emotions might rely heavily on the linguistic context for disambiguation.

While this study recorded 36 speakers, the large standard errors reported also indicate the need to collect more data. A follow-up perceptual study is also necessitated for better grouping of the phonetic parameters.

Table 1: Mean and standard error for open quotient, spectral tilt and fundamental frequency separated by speaker, gender and love

			N	Mean	
Gender	Love Type		Statistic	Statistic	Std. Error
F	Friendship	Amplitude of First Harmonic (dB)	22	7.7748	3.06398
		Amplitude of Second Harmonic (dB)	22	.6202	2.90118
		Amplitude of Third formant harmonic (dB)	22	-11.6520	3.36912
		Open Quotient (dB)	22	7.1546	1.18007
		Spectral Tilt (dB)	22	19.4269	2.00000
		Fundamental Frequency (Hz)	22	249.9578	12.62689
		Valid N	22		
	Parental	Amplitude of First Harmonic (dB)	22	10.3792	3.21943
		Amplitude of Second Harmonic (dB)	22	1.8363	3.11361
		Amplitude of Third formant harmonic (dB)	22	-11.2941	3.31325
		Open Quotient (dB)	22	8.5428	1.11083
		Spectral Tilt (dB)	22	21.6734	1.84357
		Fundamental Frequency (Hz)	22	267.9350	5.34768
		Valid N	22		
	Patriotic	Amplitude of First Harmonic (dB)	21	10.1623	3.21824
		Amplitude of Second Harmonic (dB)	21	3.0885	3.32824
		Amplitude of Third formant harmonic (dB)	21	-6.7961	2.76905
		Open Quotient (dB)	21	7.0740	1.34098
		Spectral Tilt (dB)	21	16.9584	2.06350
		Fundamental Frequency (Hz)	21	271.0073	7.84956
		Valid N	21		
	Romantic	Amplitude of First Harmonic (dB)	22	9.7109	3.26511
		Amplitude of Second Harmonic (dB)	22	3.4486	3.40327
		Amplitude of Third formant harmonic (dB)	22	-12.5868	3.42710
		Open Quotient (dB)	22	6.2622	1.23134
		Spectral Tilt (dB)	22	22.2977	1.53212
		Fundamental Frequency (Hz)	22	252.8371	5.77815
		Valid N	22		
M	Friendship	Amplitude of First Harmonic (dB)	11	8.8404	4.67546
		Amplitude of Second Harmonic (dB)	11	14.6142	4.44134
		Amplitude of Third formant harmonic (dB)	10	-4.8510	4.98984
		Open Quotient (dB)	11	-5.7738	1.75678
		Spectral Tilt (dB)	11	13.5269	2.30961
		Fundamental Frequency (Hz)	11	143.1311	7.62567
		Valid N	10		
	Parental	Amplitude of First Harmonic (dB)	13	7.1773	6.31387
		Amplitude of Second Harmonic (dB)	13	9.0268	6.65422
		Amplitude of Third formant harmonic (dB)	12	-11.1485	5.59509
		Open Quotient (dB)	13	-1.8496	1.47759
		Spectral Tilt (dB)	13	18.4812	2.45584
		Fundamental Frequency (Hz)	13	147.8305	9.42731
		Valid N	12		
	Patriotic	Amplitude of First Harmonic (dB)	12	13.1695	3.65022
		Amplitude of Second Harmonic (dB)	12	13.6551	3.26649
		Amplitude of Third formant harmonic (dB)	11	-1.0765	4.09599
		Open Quotient (dB)	12	-.4855	1.81027
		Spectral Tilt (dB)	12	13.9384	2.01819
		Fundamental Frequency (Hz)	12	193.5468	13.71375
		Valid N	11		
	Romantic	Amplitude of First Harmonic (dB)	13	10.8408	3.66021
		Amplitude of Second Harmonic (dB)	13	13.4728	3.92449
		Amplitude of Third formant harmonic (dB)	12	-4.7111	3.89301
		Open Quotient (dB)	13	-2.6322	.97555
		Spectral Tilt (dB)	13	15.6110	2.45630
		Fundamental Frequency (Hz)	13	140.6875	8.74494
		Valid N	12		

5. REFERENCES

[1] Bacharowski, J. 1999. Vocal Expression And Perception Of Emotion. *Current Directions in Psychological Science,* 53-57

[2] Banse, R. & Scherer, K. R. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology,* 70(3), 614-636.

[3] Cowie, R., & Cornelius, R. R. 2003. Describing the emotional states that are expressed in speech. *Speech Communication*, 5–32.

[4] Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., & Shibuya, Y. 2006. Acoustic and Articulatory characteristics of sad speech: Comparison of American English and Japanese. *Phonetica*, 63:1-25.

[5] Gobl, C., & Chasaide, A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, Ireland, 40(1-2), 189-212.

[6] Grichkovtsova, I., Morel, M., and Lacheret, A. 2011. The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, France, 414-429.

[7] Hanson, H. & Stevens, K. 1995. Vocal fold physiology: Voice quality control. In, Fujimura, O., Hirano, M., & Vocal Fold Physiology Conference, San Diego, Calif: Singular Pub. Group.

[8] Juslin, P. N., & Laukka, P. (2001*).* Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1, 381-412.

[9] Menezes, C. 2004. Changes in phrasing in semi-spontaneous emotional speech: Articulatory evidences. *Journal of the Phonetic Society of Japan*, 8(1), 45-59.

[10] Menezes, C., Pardo, B., Erickson, D. & Fujimura, O. 2003. Changes in syllable magnitude and timing due to repeated correction. Special Issue of *Speech Communication*, 40, issue 1-2, 71-85.

[11] Menezes, C., Erickson, D., & Han, J. 2012. Cross-Linguistic cross-modality perception of English sad and happy speech. *Proc. of the 6th International conference of Speech Prosody*, Shangai, v2, 649-652.

[12]Menezes, C., Erickson, D. & Franks, C. 2010. Comparison between linguistic and affective perception of sad and happy – A cross-linguistic study. *In Proc. of the 5th International conference of Speech Prosody*, Chicago.

[13]Panasenko, N. 2012. Linguistic Markers of Emotional Concept. *Love in Literary Texts*. *10*(4), 1067-1084.

[14] Pittam, J. & Scherer, K. 1993. *Vocal Expression and Communication of Emotion*. Handbook of Emotions, 185-197

[15] Scherer, Klaus R., 2003. Vocal communication of emotion: a review of research paradigms. *Speech Communication*. 40, 227–256.

[16] Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A., Seppānen, T., Väyrynen, E. & Airas, M. 2006 Emotions in [a]: A perceptual and acoustic study. *Logopedics Phoniatrics Vocology*, Vol. 31, No. 1, 43-48,

[17] Wierzbicka, A. 1999. Emotional universals. *Language Design*, 2, 23-69