Jitter measurements for Performance enhancement in the Service Sector

Agnes Jacob, Mythili.P
Division of Electronics
Cochin University
Kochi, Kerala, India.
nirag_2007@yahoo.co.in, mythili@cusat.ac.in

Abstract— **In a leading service economy like India, services lie at the very center of economic activity. Competitive organizations now look not only at the skills and knowledge, but also at the behavior required by an employee to be successful on the job. Emotionally competent employees can effectively deal with occupational stress and maintain psychological well-being. This study explores the scope of the first two formants and jitter to assess seven common emotional states present in the natural speech in English. The k-means method was used to classify emotional speech as neutral, happy, surprised, angry, disgusted and sad. The accuracy of classification obtained using raw jitter was more than 65 percent for happy and sad but less accurate for the others. The overall classification accuracy was 72% in the case of preprocessed jitter. The experimental study was done on 1664 English utterances of 6 females. This is a simple, interesting and more proactive method for employees from varied backgrounds to become aware of their own communication styles as well as that of their colleagues' and customers and is therefore socially beneficial. It is a cheap method also as it requires only a computer. Since knowledge of sophisticated software or signal processing is not necessary, it is easy to analyze.**

Keywords- services; communication style; formants; jitter

I. INTRODUCTION

Service based organizations include entertainment, education, health care, travel, social services and even personal services like restaurants. Recently, the realization of the services management that it is five times cheaper to retain customers than to attract new ones has increased emphasis on serving the customer. Often a customer's perception of service is affected not only by the service delivered to him, but also on the services delivered to other customers and based on how the employees interact with each other. Friendliness, courtesy and responsiveness directed towards the customers all require large amount of emotional labor from the front line staff. Emotional labor refers to the efforts to show emotions that may not be genuinely felt but must be displayed in order to express organizationally desired emotion during interpersonal transaction [1]. Any flaw in the interpersonal relationship between the service provider and the customer triggers a negative attitude towards the organization. In this context, self awareness becomes necessary to understand the potential friction at all levels and to reduce the extent of emotional labor on the job. The increasing of EI (emotional intelligence) skills (empathy, impulse control) necessary for successful job performance can help workers to deal more effectively with their feelings, and thus directly decrease the level of job stress and indirectly protect their health [2]. Emotional competence and interpersonal competence in turn are two key components of managerial competence.

Oral communication which is the main mode of employee-customer interaction in such service based organizations is very challenging and the right skills needs to be enhanced by appropriate training The research work focused in this paper attempts to assess emotions from female speech in English using minimum level of signal processing. The main parameters explored are the first two formants and the jitter. As the number of women entering the workforce has increased and more women have advanced to higher positions with greater responsibility and prestige, gender issues have become increasingly salient in the workplace [3]. Professional women who express anger may experience a decrease rather than an increase in their status since even now women evoke negative responses from other people if they fail to conform to the prescriptive stereotype. Hence it would be helpful to assess and identify one's own communication style and the most prevalent emotions in one's speech.

Formants refer to the resonant frequencies of the vocal tract and have been useful in speech analysis [4]. The perception of vowels in isolation without co articulation effects of neighboring phones is based on their steady state spectra, usually interpreted in terms of the location of F1-F3 (formants 1-3). In the normal case, virtually all vowels can be identified based on F1-F2 alone. Jitter is the perturbation in pitch. A study conducted by Lieberman and Michaels (1962) on the variation in the recognition rate due to a smoothening of the pitch contours has been reported in [5].The drop in recognition rate introduced by the smoothing of the contours was attributed to the presence of micro-perturbations known as jitter in some expressions. Those micro-perturbations of the F0 contours allowed differentiating some expressions of joy and fear, which were mistaken only when the resynthesized contours were smoothed. In 2008, it was also reported that one can extract various voice cues of relevance to speech emotion including fundamental frequency, speech rate, pauses, voice intensity, voice onset time, jitter or pitch perturbations[6].The local jitter can be used as a measure of voice quality; it is the most common jitter measurement and is usually expressed as a percentage. Local jitter is the average

absolute difference between consecutive periods, divided by the average period in waveform-matching; the duration of a period is determined by looking for best matching wave shapes. The jitter value is a measure of the irregularity of a quasiperiodic signal and is a good indicator of the presence of pathologies in the larynx which affect the voice quality [7].

II. METHODOLOGY

The speech corpus consists mainly of five vowel sounds in isolation or other words including these sounds [8].These were recorded using a high quality microphone from the utterances of six female speakers expressing seven induced emotions taken one at a time. These were segmented, labeled and stored as wave files with a sampling frequency of 16 kHz. Both the formant and jitter analysis were performed using the Praat software which has been reported to work well on long sustained vowels [9]. The waveform-matching method used here for jitter measurements averages away much of the influence of additive noise. First four formant values were noted down for each wave file. The four formants were analyzed in 2 pairs; the first one consisting of F1 and F2 and the second pair consisting of F3 and F4. All the values of both formants and jitter were first tabulated.

III. RESULTS AND DISCUSSIONS

Since F1 and F2 showed better ability to discriminate emotions, only those results are presented and discussed here for sake of relevance and brevity. From the figures 1 and 2, it can generally be concluded that the negative valence emotions have higher formant frequencies than the positive valence emotions with disgust, fear and angry having higher values for both F1 and F2. Sad has the least value for F1 alone.

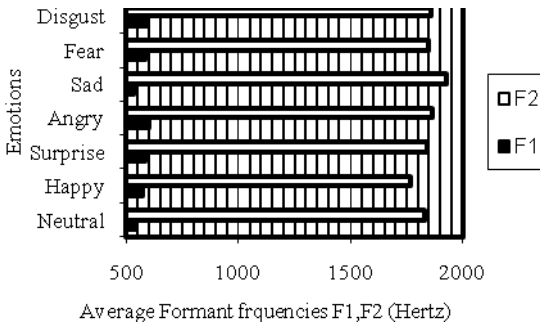

Fig.1 Comparison of first and second formant frequencies for seven emotions

There is also considerable difference between the first formant frequency value and the second formant frequency value for each of the seven emotions. In Fig.2, the scatter plot of the first two formant frequencies for five different utterances under the seven different emotions shows the scope of F1 and F2 to discriminate between various sounds rather than between various emotions. The figure also indicates the mix up of the first two formant values for certain emotions under each utterance.

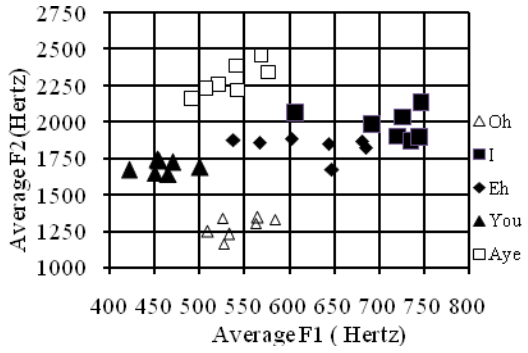

Fig. 2 Average F2 versus F1 for various vowel utterances under 7 emotions

Classification of randomly selected formant values using the k-means classifier gave a maximum recognition accuracy of around 40% only. However, by a similar approach the formants were found efficient in discriminating between utterances as mentioned in [4]. The k-means Classifier based on F1-F2 values gives 100% accuracy in results in grouping a, o, u. Regarding Eh, there was 86% accuracy ,the remaining 14% inaccuracy due to confusion with you. The highest confusion was between I and Eh; with an accuracy of only 57% in the recognition of I. In the other 43% cases there was confusion with Eh. Next it was decided to look into the possibilities of using jitter for emotion assessment. The following parameters presented in Table. I were derived from the raw values of jitter measurement

TABLE I. STATISTICAL PARAMETERS OF JITTER

Emotions	Jitter Values			
	Mean	*Std.devn*	*Max*	*Min*
Neutral	0.0215	0.0076	0.0387	0.01
Surprise	0.0236	0.008	0.039	0.0097
Happy	0.0220	0.0052	0.031	0.0127
Angry	0.0269	0.0096	0.0437	0.01
Sad	0.0152	0.007	0.035	0.0057
Fear	0.0296	0.0135	0.0657	0.011
Disgust	0.0214	0.0088	0.038	0.007

The highest mean jitter as well as the maximum jitter occurs under the fear emotion. So also the least mean jitter as well as the least jitter value occurs for sad as seen in Fig .3. The highest standard deviation from the mean occurs for fear, while the lowest SD from the mean is for Happiness. Statistical analysis was also performed on the data to ensure that the jitter values for different emotions were significantly different. One way standard ANOVA was performed on the jitter data.

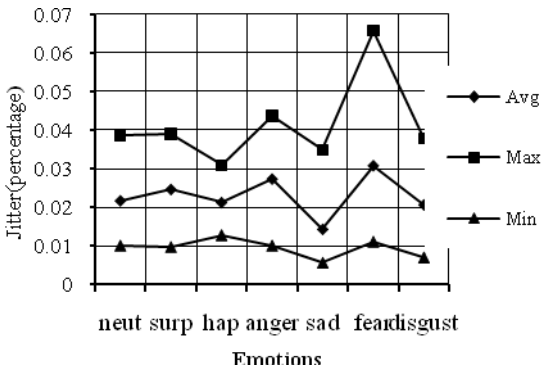

Fig.3 Average, maximum, minimum values of jitter for various emotions

In Fig.4, unlike the case of formants we observe distinct clusters under different emotions indicating that jitter is a relevant parameter for classification of different emotions.

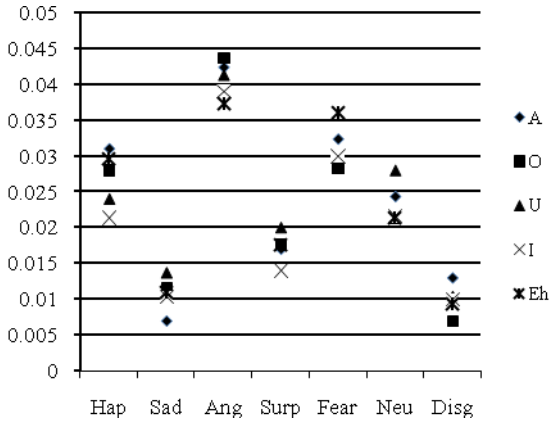

Fig .4 Plot showing distinct jitter based clusters for different emotions

Classification of jitter values using the k-means algorithm has yielded the following emotion recognition accuracies given in Table II. The highest recognition accuracy is 75% for the sad emotion followed by happy, while the highest error in emotion recognition occurs for the neutral. 8.3% of sad utterances and the remaining 16.6% of sad utterances are confused with surprise and disgust respectively, thus totaling to 100 percent. It can also be observed under the column for sad that, 25% of disgust only has been confused with sadness. Disgust is seen to be confused with all other emotions considered here. Though the overall recognition accuracy is less for this female speech database, the recognition accuracies are higher for the happy and sad emotions than those reported using discrete wavelet transforms and Artificial Neural networks to discriminate neutral, happy, anger and sad from Malayalam utterances. [10]. In our present study too, the classification accuracy would have logically increased if we had restricted our study to fewer emotion classes. This is most evident from the fact that nearly 33% of utterances under fear have been classified as angry ones. Alternatively it was found that a deeper level of statistical analysis yields better classification results. Classification of the raw jitter values of happiness, sadness, anger and surprise yielded more than a chance (greater than 50%) recognition of emotions.

TABLE II: CONFUSION MATRIX OF JITTER VALUES

*Emo	Happy	Sad	Anger	Surp	Fear	Neut	Disg
Happy	66.7%	0	8.3%	8.30%	0	8.30%	8.30%
Sad	0	75%	0	8.30%	0	0	16.6%
Anger	25%	0	58%	0	8.3%	0	8.3%
Surp	8.3%	0	0	58%	0	8.30%	25.0%
Fear	33.3%	0	33.3%	0	25%	0%	8.3%
Neut	17%	0	25%	33.0%	0	17%	8.0%
Disg	17%	25%	16.6%	8%	0	0	33%

*Emo-emotions; surp-surprise; neut-neutral; disg-disgust.

Table. III gives the confusion matrix wherein standard ANOVA had been performed on the jitter values to ensure significant difference between the jitter values of various emotion classes. As only those values with more than two star significance were given to the k- means classifier, better classification results have been obtained in this case. Thus jitter is also showing the potential for emotion classification as it is able to categorize correctly, happiness, sadness, surprise and to a certain extent fear.

TABLE III: CONFUSION MATRIX OF PREPROCESSED JITTER VALUES

Emo	Happy	Sad	Anger	Surp	Fear	Neut	Disg
Happy	100%	0	0	0	0	0	0
Sad	0	78%	0	0	0	0	22%
Anger	56%	0	33%	0	11%	0	0
Surp	0	0	0	89%	0	0	11%
Fear	22%	0	11%	0	67%	0	0
Neut	33%	0	0	0	0	67%	0
Disg	0	11%	0	22%	0	0	67%

IV. CONCLUSION

Greater accuracies in emotion recognition were obtained when the data belonging to the different emotion classes were statistically verified as significantly different. But often such a pre-processing of data is time consuming and robs the simplicity of the procedure. This study has used the K means method to classify emotions based on jitter and formants, though the performance is much less for the first two formant frequencies. But F1 F2 are efficient to classify utterances. The approach used here is totally non invasive, affordable technique and can be included as a part of the sensitivity training conducted by the human resources departments in the organization. This activity is highly beneficial in the case of high contact employees whose physical presence near the customer and interactions extend over a long period of time. The main aim of this activity is to enhance the functional quality of the employees and society as a whole through improved level of emotion consciousness. The benefits could be verified after applying the method and observing the changes in employee behavior and competencies.

REFERENCES

[1] Mann, S. "Emotion at work: To what extent are we expressing, suppressing, or faking it?" European Journal of Work and Organizational Psychology, 8, 1999.pp .347–369.

[2] Oginska. Bulik, Nina, "Emotional intelligence in the workplace: Exploring its effects on occupational stress and health outcomes in human service workers". International Journal of Occupational Medicine & Environmental Health, 28(2), 2005.pp.167-175.

[3] Vijai N Giri, Gender Role in Communication Style. IIT Kharagpur. Concept Publishing House, New Delhi.11005

[4] Douglas O' Shaughnessy, Speech Communication :Human and machine. Addison Wesley Publication, 1987.

[5] Tange Banziger, Klaus R Scherer, "The role of intonation in emotional expressions." Elsevier Speech Communication, vol 46, pp.252-265, 2008.

[6] Patrik N. Juslin and Klaus R. Scherer, Scholarpedia, 2008. 3(10):4240

[7] D´arcio G. Silva,1 Lu´ıs C. Oliveira and M´ario Andrea**, "**Jitter estimation algorithms for detection of pathological Voices." Hindawi publishing corporation, EURASIP Journal on advances in signal processing,Volume 2009, Article ID 567875, 9 pages, doi:10.1155/ 2009/567875.

[8] Daniel Jones, English Pronouncing Dictionary, Fourteenth Edition, Cambridge University Press ,2004.

[9] P.Boersma and D.Weenink, "Praat: doing phonetics by Computer (Version 4.6.09)," 2005[Computer Program],http:/ /www.Praat. org/;

[10] Firoz Shah, A.Raji Sukumar , A and Babu Anto P "Discrete wavelet transforms and artificial neural networks for speech emotion recognition." International Journal of Computer Theory and Engineering , Vol 2, No3, June 2010.