ACOUSTIC ANALYSIS OF MANDARIN SPEECH IN A MIXED EMOTION OF PLEASURE AND DISGUST

Mengjie Gong[1] and Wentao Gu[1,2]

[1]School of Chinese Language and Literature, Nanjing Normal University, Nanjing, China
[2]School of Communication Sciences and Disorders, McGill University, Montreal, Canada
gmjnew@qq.com, wtgu@njnu.edu.cn

ABSTRACT

It is controversial whether people can experience a mixture of two emotions with opposite valences. Although a number of psychological and physiological studies supported the existence of mixed emotions, acoustic-phonetic characteristics of speech in mixed emotions are yet to be explored. This study investigated a mixed emotion composed of pleasure and disgust. Statistical analysis of acoustic parameters showed that the main difference among positive (pleasure), negative (disgust) and mixed emotions existed in fundamental frequency parameters, harmonic differences, and Harmonic-to-Noise Ratio (HNR), while the intensity parameters, Jitter and Shimmer were not significantly different. Clustering analysis further showed that the three types of emotions could be classified more effectively in terms of the acoustic parameters showing significant difference.

Keywords: mixed emotion; Mandarin; non-parametric statistical analysis; clustering analysis

1. INTRODUCTION

In terms of emotional valence, emotion can be positive or negative. Although a number of studies investigated the acoustic manifestations of speech conveying positive or negative emotions [1, 2], little has been explored for mixed emotions that incorporate multiple emotional states simultaneously.

In the literature, *mixed emotion* has been defined in two distinctive ways. Larsen et al. [3] proposed that emotional experience can be characterized by the concurrent activation of two emotions that are typically opposite in valence, e.g., happiness and sadness. In contrast, Grühn et al. [4] suggested that mixed emotion is an emotional state that lies between two emotions with opposite valences, and values closer to zero indicate a greater emotional complexity. So far, most studies on mixed emotions are based on the first definition, i.e., positive and negative emotions are concurrently elicited. Larsen and McGraw [5] demonstrated the robustness of this mixed emotion as a genuine and distinct emotional experience. Moreover, previous studies have elucidated the psychological and physiological differences between a mixed emotion and a pure positive or negative emotion [6, 7].

However, up to now few research has looked into the acoustic manifestations of the speech conveying mixed emotions. Thus, the present study aimed to investigate prosodic and voice quality characteristics of Mandarin utterances conveying a mixed emotion of “pleasure” (positive valence) and “disgust” (negative valence), in comparison with those conveying a pure pleasure (positive) or disgust (negative) emotion.

2. METHODS

2.1. Materials

A film-watching paradigm was employed to elicit emotions. For this purpose, we selected nine film clips from the database created by Samson et al. [8], and had them evaluated by 27 subjects (age: *M*=23.9, *SD*=1.65) for the scores of pleasure and disgust on the respective five-point Likert scale. These film clips were divided into three blocks, and within each block three clips presented in a random order were evaluated by nine subjects.

A mixed emotion coefficient (MF) [9, 10], defined as the minimum of the pleasure and disgust scores, indicates the intensity of the mixed emotion. For example, if pleasure is scored 3 and disgust is scored 4, MF will be 3. When one of the pleasure and disgust scores is 1, the emotion is regarded as a pure positive or negative one. All MF scores above 1 for each clip are shown in Fig. 1, where each data point indicates a subject. Finally, two clips (“hit” and “downstairs”) with the highest number of MF scores above 1 (hence most likely to elicit a mixed emotion) were selected.

Also, we designed twelve 7-syllable Mandarin target sentences without any emotional words. Naturalness and emotion-neutrality of the sentences were validated by 15 native speakers of Mandarin.

2.2. Participants

Twenty-nine native speakers of Mandarin (12M, 17F; age: *M*=24.3, *SD*=1.5) without any history of hearing, speech, and emotion impairments were recruited.

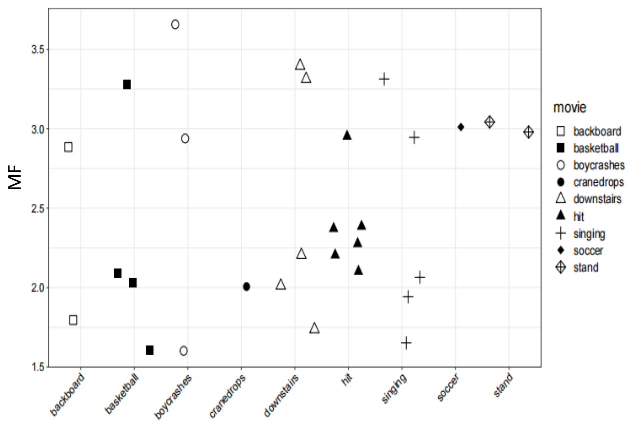

Figure 1: MF scores of film clips.

2.3. Procedure

After watching each film clip, participants scored pleasure and disgust on the respective 5-point Likert scale based on their visual impression of the film clip. Then, they were asked to say 6 sentences (a random selection from the 12 target sentences) with the emotion they just perceived from the film clip. Before watching the next clip, there was a brief break (at least 1.5 min) such that the emotional states of the participants could return to their baseline level.

2.4. Speech data

A total of 6*2*29 = 348 utterances were recorded. Among them, 150, 84, and 108 utterances were associated respectively with the mixed, positive, and negative emotions, based on the pleasure and disgust scores labelled by participants on the film clips. Only 6 utterances were neutral (i.e., both pleasure and disgust were scored 1), which were excluded in subsequent analysis due to lack of tokens.

2.5. Data analysis

The commonly-used acoustic parameters, including fundamental frequency (F0) contour, intensity (INT) contour, H1–H2, H1–A1, H1–A2, Harmonic-to-Noise Ratio (HNR), Jitter and Shimmer, were extracted from each utterance using the Praat toolkit [11]. For F0 and INT contours, the mean, standard deviation, minimum, maximum, and range were further calculated.

All these 16 parameters were statistically analysed with R [12]. Because none of them passed the Shapiro-Wilk normality test, non-parametric Scheirer-Ray-Hare tests were then conducted, with speaker gender and emotion as two independent factors. However, results showed that there was no significant interaction effect between the two factors. Thus, to simplify analysis, non-parametric Kruskal-Wallis tests were conducted instead, with emotion (Pos, Neg, Mix) as the sole independent factor, without taking speaker gender into account.

Table 1: Results of Kruskal-Wallis test on all 16 parameters.

	Median value			χ^2	df	*p*-value
	Pos	Neg	Mix			
F0mean	151.323	209.186	212.136	7.923	2	0.019*
F0std	25.868	42.740	29.812	13.593	2	0.001**
F0min	105.574	107.868	136.946	9.446	2	0.009**
F0max	207.956	269.376	264.974	5.714	2	0.057
F0range	93.604	165.162	112.933	13.622	2	0.001**
INTMean	66.229	65.747	66.552	3.125	2	0.210
INTStd	8.350	8.495	8.109	1.094	2	0.579
INTMin	39.709	38.403	39.549	1.138	2	0.566
INTMax	77.847	78.095	77.412	2.629	2	0.269
INTRange	38.765	39.692	37.367	3.514	2	0.173
Jitter	0.026	0.023	0.020	5.710	2	0.058
Shimmer	0.094	0.071	0.084	4.590	2	0.101
H1-H2	5.072	6.317	7.563	13.603	2	0.001**
H1-A1	17.985	18.492	20.107	7.875	2	0.020*
H1-A2	20.546	17.839	21.814	7.209	2	0.030*
HNR	14.731	18.796	17.917	7.713	2	0.021*

Table 2: Results of post-hoc tests on eight parameters.

	Pos vs Mix	Mix vs Neg	Pos vs Neg
F0mean	*		*
F0std		**	**
F0min	*	*	
F0range		**	**
H1-H2	**		
H1-A1	*		
H1-A2		*	
HNR			*

Note: * $p < 0.05$; ** $p < 0.01$.

3. RESULTS

3.1. Non-parametric tests

Table 1 shows the results of Kruskal-Wallis tests on all 16 acoustic parameters, among which eight parameters exhibited significant effects of emotion. For these eight parameters, Fig. 2 shows the boxplot for three emotions, and Table 2 shows the results of post-hoc tests with Bonferroni correction using the *PMCMRplus* package [13]. The significant results are summarized as follows:

F0mean: Mix > Pos ($p = 0.049$), and Neg > Pos ($p = 0.023$).

F0std: Neg > Mix ($p = 0.005$), and Neg > Pos ($p = 0.004$).

F0range: Neg > Mix ($p = 0.004$), and Neg > Pos ($p = 0.005$).

F0min: Mix > Neg ($p = 0.025$), and Mix > Pos ($p = 0.044$).

H1–H2: Mix > Pos ($p = 0.001$).

H1–A1: Mix > Pos ($p = 0.030$).

H1–A2: Mix > Neg ($p = 0.020$).

HNR: Neg > Pos ($p = 0.026$).

3.2. Clustering analysis

Using the Partitioning Around Medoids (PAM) clustering algorithm which is a robust alternative to K-means, we further conducted clustering analysis with the above eight parameters.

Before PAM clustering, the pamk() function in the *fpc* package [14] was run to find the optimal number of clusters, which turned out to be three, coinciding exactly with the number of emotion types here. The data were then clustered using the pam() function in the *cluster* package [15]. The first and second principal components explained 72.94% of variance. The clustering results based on the two principal components are shown in Fig. 3, where categories 1 and 2 have the most conspicuous overlap, categories 1 and 3 overlap to a much lesser extent, while categories 2 and 3 overlap the least.

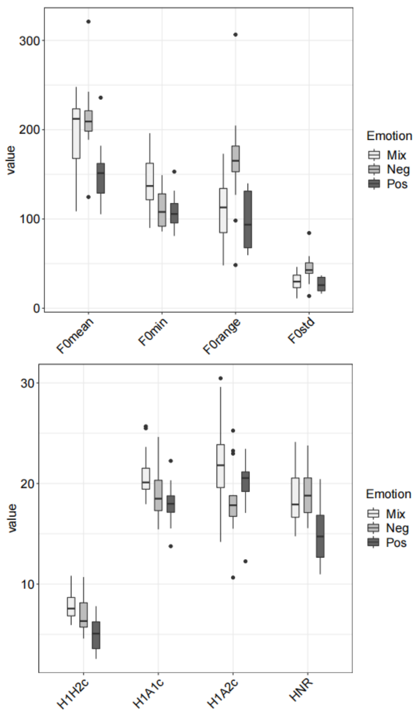

Figure 2: Comparison of eight parameters in three emotions.

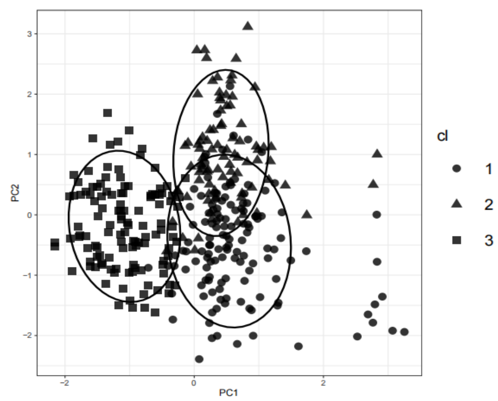

Figure 3: Results of PAM clustering analysis.

Table 3 shows the contingency table for clustering. For one thing, categories 1 and 2 correspond mainly to negative and mixed emotions, respectively, while category 3 corresponds almost equally to positive and mixed emotions. For another, positive and negative emotions are clustered mainly into categories 3 and 1, respectively, while the mixed emotion are distributed almost evenly in all three categories.

Table 3: Contingency table for clustering analysis.

	Emotion		
	Pos	Mix	Neg
Category 1	15	41	78
Category 2	11	56	20
Category 3	58	53	10

4. DISCUSSION AND CONCLUSION

To our knowledge, this study is the first attempt to look into the acoustic characteristics of speech in a mixed emotion. Results showed that the Mandarin utterances conveying a mixed emotion of pleasure and disgust differed significantly from those conveying a pure pleasure or disgust emotion in seven acoustic parameters, including F0 features (F0mean, F0std, F0min, F0range) and harmonic differences (H1–H2, H1–A1, H1–A2), while HNR differed significantly only between pleasure and disgust emotions. Furthermore, clustering analysis verified that with these eight parameters the speech data could be approximately clustered into three categories.

There are still some limitations in this preliminary study. First, we examined a mixed emotion of pleasure and disgust, since previous studies had demonstrated the reliability of this emotion [10, 16]. Nevertheless, future research needs to address whether the conclusions here can be generalized to other mixed emotions [17]. Second, film clip elicitation, as a standard experimental paradigm to stimulate pure or mixed emotions [10], is easy to standardize and can elicit highly ecologically valid emotions [18]. However, film clips may differ in various potential characteristics, may induce multiple emotions sequentially rather than simultaneously, and subjects in this paradigm may become passive observers of events rather than active participants. Therefore, our future study needs also to examine whether the findings here can be generalized to other emotional situations.

5. ACKNOWLEDGEMENTS

This work is supported by the National Social Science Fund of China (13&ZD189).

6. REFERENCES

[1] Colombetti, G. 2005. Appraising valence. Journal of Consciousness Studies 12(8-10), 103-126.
[2] Kensinger, E. 2004. Remembering emotional experiences: The contribution of valence and arousal. Reviews in the Neurosciences 15(4), 241-252.
[3] Larsen, J.T., McGraw, A.P., and Cacioppo, J. T. 2001. Can people feel happy and sad at the same time? J. Pers. Soc. Psychol. 81, 684-696.
[4] Grühn, D., Lumley, M.A., Diehl, M., et al. 2013. Time-based indicators of emotional complexity: Interrelations and correlates. Emotion 13, 226-237.
[5] Larsen, J.T., and McGraw, A.P. 2014. Further evidence for mixed emotions. J. Pers. Soc. Psychol. 100, 1095-1110.
[6] Kreibig, S.D., Samson, A.C. and Gross, J.J. 2013. The psychophysiology of mixed emotional states. Psychophysiology 50(8), 799-811.
[7] Kreibig, S.D., Samson A.C. and Gross, J.J. 2015. The psychophysiology of mixed emotional states: Internal and external replicability analysis of a direct replication study. Psychophysiology 52(7), 873-886.
[8] Samson, A.C., Kreibig, S.D., Soderstrom, B., et al. 2016. Eliciting positive, negative and mixed emotional states: A film library for affective scientists. Cognition and Emotion 30(5), 827-856.
[9] Schimmack, U. 2001. Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? Cognition and Emotion 15, 81-97.
[10] Hemenover, S.H., and Schimmack, U. 2007. That's disgusting!..., but very amusing: Mixed feelings of amusement and disgust. Cognition and Emotion 21, 1102-1113.
[11] Boersma, P. and Weenink, D. 2022. Praat: Doing phonetics by computer (Version 6.3) [Computer program]. Retrieved 15 November 2022. Available from http://www.praat.org/
[12] R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[13] Pohlert, T. 2022. PMCMRplus: Calculate pairwise multiple comparisons of mean rank sums extended. R package version 1.9.6. https://CRAN.R-project.org/package=PMCMRplus
[14] Christian, H. 2020. fpc: Flexible procedures for clustering. R package version 2.2-9. https://CRAN.R-project.org/package=fpc
[15] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. 2021. cluster: Cluster analysis basics and extensions. R package version 2.1.2.
[16] Oppliger, P.A., & Zillmann, D. 1997. Disgust in humor: Its appeal to adolescents. Humor 10, 421-437.
[17] Harter, S., & Buddin, B.J. 1987. Children's understanding of the simultaneity of two emotions: A five-stage developmental acquisition sequence. Developmental Psychology 23, 388-399.
[18] Rottenberg, J., Ray, R., & Gross, J. 2008. Emotion elicitation using film clips. In J. A. Coan & J. J. B. Allen (Eds.), The Handbook of Emotion Elicitation and Assessment. New York: Oxford University Press, 9-28.