An Analysis of the Speech Under Stress Using the Two-Mass Vocal Fold Model

Xiao Yao, Takatoshi Jitsuhiro, Chiyomi Miyajima,
Norihide Kitaoka, Kazuya Takeda

Abstract We focus on the glottal source of speech production, which is essential for understanding the behavior of vocal fold when speech is produced under psychological stress. A spectral flatness measure (SFM) is introduced, as a useful tool, to evaluate stress levels in speech. Further, the relationship between the physical parameters of the two-mass vocal fold model and the proposed stress level measurement is established. The physical parameters of two-mass model are examined and analyzed comparing with measurements in order to estimate the state of vocal folds in people experiencing stress in the future. In this paper, experiments are performed using stressed speech gathered from real telephone conversations to evaluate the stress level measurement.Results show that the SFM can detect stress and can be used as a measurement for differentiating stressed from neutral speech. Furthermore, the changes in physical parameters can be analyzed to understand the behavior of vocal folds when stress occurs.

Xiao Yao
Graduate School of Information Science, Nagoya University, Aichi, Japan, e-mail: `xiao.yao@g.sp.m.is.nagoya-u.ac.jp`

Takatoshi Jitsuhiro
Department of Media Informatics, Aichi University of Technology, Gamagori Japan, e-mail: `jitsuhiro@aut.ac.jp`

Chiyomi Miyajima
Graduate School of Information Science, Nagoya University, Aichi, Japan, e-mail: `miyajima@nagoya-u.jp`

Norihide Kitaoka
Graduate School of Information Science, Nagoya University, Aichi, Japan, e-mail: `kitaoka@nagoya-u.jp`

Kazuya Takeda
Graduate School of Information Science, Nagoya University, Aichi, Japan, e-mail: `kazuya.takeda@nagoya-u.jp`

1 Introduction

It has recently become much more important to estimate someone's mental condition, especially from speech. Particularly for call center systems, detection techniques applied to caller speech can significantly improve customer service. A study of speech under stress is needed, in order to improve recognition of the mental state of people, and to understand the context in which the speaker is communicating.

Many scholars have devoted their researches to the field of analysis of stress in speech. Stress is a psycho-physiological state characterized by subjective strain, dysfunctional physiological activity, and deterioration of performance [1]. Some external factors and internal factors may induce stress, which are likely to be detrimental to the performance of communication equipment and systems with vocal interfaces [2]. The performance of a speech recognition algorithm is significantly challenged when speech is produced in stressful environments. The influence of the Lombard effect on speech recognition has been focused on in [3], and in some special environments workload task stress has been proven to have a significant impact on the performance of speech recognition system [4]. Also some work has been done from a linear speech production model for the classification of stressed speech [5].

Our work mainly concentrates on the analysis of stressed speech based on a speech production model instead of on observed speech features, in order to gain a deeper comprehension of the working mechanisms of vocal folds responsible for different speaking styles. Our final target is to simulate several speaking style speech by the two-mass model for detecting stress speech. For this purpose, we will explore the properties of the underlying physical speech production system, and search for essential factors related to stress. The glottal flow, which is the source of speech, is mainly examined when making an attempt to explain changes in speech production under stress. As a result of this study, the characteristics of glottal flow can be related to physical parameters, and an explanation of how the two-mass model applies to real speech can be made.

In this paper, our recent works are described that the physical parameters of two-mass model are examined and analysed comparing with stress measurement. a measurement is introduced to evaluate the presence of stress in speech in the real world. We concentrate on analysis of the ability of the model to detect the presence of stress by exploring the relationship between the stress level measurement and parameters of two-mass model and then variation in stiffness is studied to estimate the behaviour of the vocal folds. The paper is structured as follows: In section 2, the stress level measurement used to detect the stress in this paper is described. In section 3, simulation of speech using two-mass model is described and used to find the relationship between the stress level measurement and stiffness parameters, and experimental results are shown and analysed to prove that such a relationship exists. Finally, in section 4 we draw our conclusions.

2 Measuring stress using glottal source

2.1 Spectral flatness of the glottal flow

The glottal flow can be estimated by the inverse filter method from speech pressure waveforms using the iterative adaptive inverse filter method (IAIF). To evaluate the stress from glottal flow, the spectral flatness measure (SFM) can be applied. Spectral flatness is a measurement to characterize an audio spectrum, which defined by dividing the geometric mean by the arithmetic mean of the power spectrum:

$$SFM = \frac{\sqrt[N]{\prod_{n=0}^{N-1} S(n)}}{\frac{1}{N}\sum_{n=0}^{N-1} S(n)}, \tag{1}$$

in which S(n) is the magnitude of the power spectrum of the nth bin. Bigger SFM indicates that the spectrum has a similar amount of power in all frequency bands, and the speech spectrum envelope would become relatively flat, like white noise. While smaller SFM shows that the spectral power is concentrated in relatively narrow bands, and its envelope would appear spiky, as is typical for tonal sound.

2.2 Evaluation of Spectral Flatness Measure

In the experiment, we used a database collected by the Fujitsu Corporation[13]. This database contains speech samples from 11 people, including 4 male and 7 female subjects. In order to simulate mental pressure resulting in psychological stress, three tasks corresponding different situations were introduced. These tasks were performed by a speaker having a telephone conversation with an operator to simulate a situation involving pressure during telephone communication. The three kinds of tasks are (A) Concentration; (B) Time pressure; and (C) Risk taking. For each speaker, there are four dialogues with different tasks.

The speech data from database is inverse filtered with 12-order LPC. The frame size is 64ms, and with 16ms for frame shift. The distributions of SFM are shown in Figure 1.

From the results of our experiment, stressed speech can be separated from neutral speech by analysing the distribution of the stress level measurement: SFM, Compared to the values for neutral speech, stressed speech result in smaller SFM values.

3 Simulation using two-mass model

Two-mass vocal fold model is proposed by Ishizaka and Flanagan to simulate the process of speech production [6]. The relationship between the physical parameters

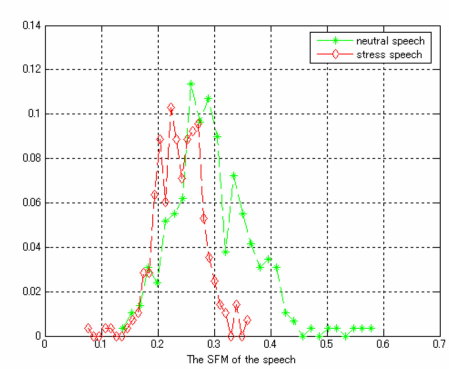


Fig. 1 Distribution of SFM for neutral and stress speech

and stress level measurement is essential for the work to analysing stressed speech using the model. Therefore, the physical parameters of two-mass model are checked and analysed comparing with stress measurement. It would provide the basis to make some assumptions from which we can estimate the shape and movement of the vocal folds when speech is produced under the stressful conditions.

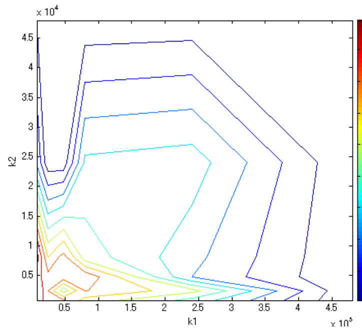
The control parameters [7] are defined as Ps, k1, k2, kc (sub-glottal pressure and stiffness), which would influence the fundamental frequency. The initial areas Ag0i have been set as 0.05cm2, and the masses and the resistance viscosity set equal to typical values proposed by Ishizaka and Flanagan.

To obtain a significant relationship between the control parameters and the proposed stress level measurement, different values of stiffness are analyzed to study the variation in the SFM. Three stiffness parameters k1, k2, kc should be taken into account. The range of for k1 is from 8000 to 480000, and k2 from 800 to 48000, and the range from 2500 to 150000 for kc. In this experiment, only glottal flow is considered, which represents the source of speech. Using a fixed set of parameters, the glottal flow is obtained to calculate the values of stress measurement.

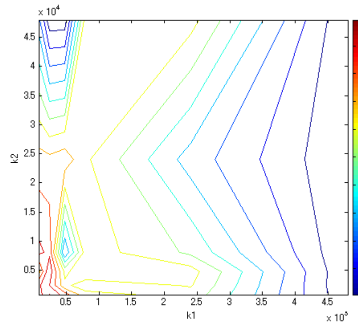
First, kc is fixed, and the variation in SFM is observed with k1 and k2. The results are shown in Figure 2.

kc is fixed as different values: (a) 2,500 (b) 7,500 (c) 25,000. When kc is fixed to a smaller value (a), SFM depends on both k1 and k2, A clear decrease in SFM is noticed as k1 and k2 become stronger, so this observation indicates that bigger values of both k1 and k2 can result in smaller values of SFM, which represents stress; When kc increases to 7,500 (b), SFM is more strongly influenced by k1 than k2; when a bigger value of kc is fixed in (c), SFM only depends on the larger k1; If kc continues increasing, a similar trend can be observed.

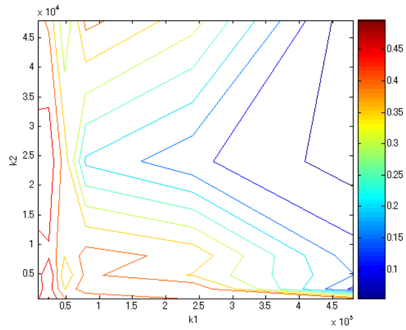
Figure 3 exhibits the relation between the SFM and each stiffness parameter respectively. The variation range of SFM with each stiffness is k1: 0.3689, k2: 0.1680, kc: 0.1376. Therefore, SFM is more strongly influenced by k1 than k2 and kc. In relation to SFM, the values of k1 and k2 are inversely proportional to SFM. The difference is that k1 is stable at first, but declines rapidly soon afterwards, while



(a) kc =2,500



(b) kc =7,500



(c) kc =25,000

Fig. 2 The relationship between the estimated stress level measurement (SFM) and stiffness parameters

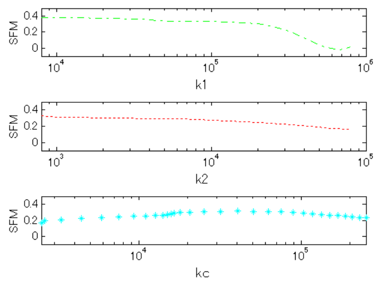


Fig. 3 The relationship between the mean of the estimated stress level (SFM) and stiffness parameters

k2 always declines steadily. The value of SFM increases with the coupling stiffness of kc at beginning, and decrease finally, which means a smaller and a larger kc are also a better indicator of stress detection. Therefore, when stress occurs, kc would normally be smaller or larger, but a bigger k1 and k2 could be obtained.

4 Conclusion

We developed a method of analyzing stress in speech by using a speech production model to explore the underlying mechanisms of vocal fold behavior. In this work, a stress level measurement method was applied. The relationship between the physical parameters and the SFM value has been established, and the physical parameters of two-mass model were checked and analyzed comparing with stress measurement. Further improvements are possible if more detailed investigation, from the view-point of vocal dynamics, is needed to build a theoretical basis of stressed speech production. More works will be focused on the extraction of physical parameters from real speech concerning vocal folds and vocal tract, and an explanatory relationship between the physical parameters and stress will be built.

5 Acknowledgements

This work has been partially supported by the ”Core Research for Evolutional Science and Technology”(CREST) project of the ”Japan Science and Technology Agency”(JST). We are very grateful to Mr. Matsuo of the Fujitsu Corporation for the use of database and for his valuable suggestions.

References

[1] Steeneken, H.J.M. and Hansen, J.H.L.: Speech Under Stress Conditions: Overview of the Effect on Speech Production and on System Performance. in Proc. ICASSP, **4**, 2079-2082 (1999).
[2] Cairns, D, Hansen, J.H.L.: Nonlinear Analysis and Detection of Speech Under Stressed Conditions. The Journal of the Acoustical Society of America, textbf96, **6**, 3392-3400 (1994).
[3] Junqua, J.C. : The Lombard reflex and its role on human listeners and automatic speech recognizers. J. Acoust. Soc. Amer., **1**, 510-524, (1993).
[4] Bard, E. G., Sotillo, C., Anderson, A. H., Thompson, H. S., and Taylor, M. M.: The DCIEM map task corpus: Spontaneous dialogue under sleep deprivation and drug treatment, Speech Commun., **20**, 71-84 (1996).
[5] Hansen, J. H. L.: Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition, Speech Commun., **20**, 151-173 (1996).
[6] Ishizaka, K., Flanagan, J.L.: Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell.Syst.Tech .Journal, **51**, 1233-1268 (1972).
[7] Teffahi, H.: A two-mass model of the vocal folds: determination of control parameters. Multimedia computing and System (2009).