Evaluation of Acoustic Correlates of Speech Under Stress for Robust Speech Recognition

John H.L. Hansen
Department of Electrical Engineering
Duke University
Durham, North Carolina 27706

1 Abstract

This paper presents results from an investigation of how speech characteristics change under varying levels of stress with specific application to improving automatic isolated-word speech recognition. In JASA-87 [2], preliminary results were presented, based on a series of probe studies which served to identify possible stress relayers in a speech recognition/communication framework. This paper presents results from subsequent evaluations which are more comprehensive and unified in addressing the problem of speech under stress [3]. Evaluation focused on five speech analysis domains: i) pitch, ii) glottal source, iii) intensity, iv) duration, and v) vocal tract shaping. Goodness-of-fit statistical tests were used to ascertain the significance of parameter variation in each domain. This paper will present results from analysis of pitch and glottal source spectrum. Findings suggest that such parameter information can be used reliably to aid in automatic isolated-word speech recognition in noisy stressful environments.

2 Introduction

Although much research has been dedicated to formulating new speech recognition techniques, little has been learned about why recognizers sometimes fail. It has been suggested that one of the overriding factors which contribute to recognition errors is talker variability between training and actual environmental conditions. Also, limited progress has been maded in addressing the issue of diverse environmental conditions for speech recognition. This is due in part to the fact that past approaches have largely been applied in tranquil environments. Studies have shown that recognition accuracy is severely reduced when speech is uttered in a noisy, stressful environment. If recognition is to be successful in such diverse environments, (e.g., pilots in aircraft cockpits, wheelchair control for the disabled, factory use for assembly lines), effects of environmental conditions such as noise and stress on voice characteristics must be taken into account.

The goal, is to improve recognition capabilities of speech produced under stressful conditions. To accomplish this, speech parameters most affected by environmental conditions must be identified. A comprehensive and unified investigation was performed which revealed new and statistically reliable acoustic correlates of speech under stress[3]. Well over 200 parameters, based on characteristics from five speech production domains (e.g., pitch, glottal source, duration, intensity, vocal-tract) were considered in the evaluation. Statistical evaluation of average, variance, and the distribution of these parameters were considered. This paper will summarize some key source excitation results from this evaluation. An in depth discussion of all the speech analysis domains can be found in [3]. Stress in this context refers to the result of factors which act on the speaker from environmental conditions (e.g., task induced workload stress, background noise as in the Lombard effect[5], etc).

3 Speech Under Stress Data Base

To improve recognition performance, speech parameters most affected by environmental conditions such as stress or noise must first identified. Previous research directed at this problem has generally been limited in scope, often suffering from one to five problems. These include: i) limited speaker populations, ii) sparse vocabularies, iii) qualitative results with little statistical confirmation, iv) limited numbers and types of speech parameters considered, and v) analysis based on simulated or actual conditions with little confirmation between the two. In order to address these issues, a comprehensive speech under stress data base was established for the purposes of stress research [3]. Table 1 presents the five domains which include: various talking styles (slow, fast, soft, loud, angry, clear, question), single and dual tracking workload stress inducing tasks, emotional speech from psychiatric analysis sessions, speech spoken in noise (Lombard effect), and subject motion-fear tasks.

SPEECH UNDER STRESS DATA BASE Department of Electrical Engineering				
Domain	*Type of Stress or Emotion*	*Number of Speakers*	*Number of Utterances*	*Source*
Psychiatric Analysis	Depression Fear, Anger Anxiety	6 Female, 2 Male	600 (present)	Emory University School of Medicine Department of Psychiatry
Talking Styles	slow, fast soft, loud angry, clear question	All Male 3 General 3 New York 3 Boston	8820 (total)	Lincoln Labs Boston, Mass. 35 aircraft communication words
Single Tracking Task	*Workload* (moderate-C50) (high-C70) Lombard	All Male	1890 (total)	Lincoln Labs Calibrated Workload Tracking Task
Dual Tracking Task	*Workload* (moderate) (high)	4 Female 4 Male	4320 (total)	Georgia Tech Acquisition tracking Compensatory tracking
Subject Motion-Fear Tasks	G-force Lombard,noise anxiety,fear	3 Female 4 Male	400 (total)	Georgia Tech Controlled Motion Noisy Environment

Table 1: The GT.JH Speech under Stress Data Base.

4 Analysis of Speech Under Stress

The database evaluation was partitioned into three areas. Analysis was first performed on *(i) speech with simulated stress* and *(ii) speech from stress inducing workload tasks or speech in noise.* Statistically significant parameters were established, and an equivalent analysis was perfromed over *(iii) speech produced under actual stress or emotion.*

CH2689-8/89/0000-0031 $1.00 © 1989 IEEE

4.1 Analysis of Pitch

Due to the structure of the data base, speaker strategies for conveying stress such as conversational context and sentence structure were ruled out for analysis. To reduce the effects of lexical stress in the analysis of pitch, a majority of the simulated stress utterances analyzed were monosyllabic. Analysis of pitch comprised five evaluation steps: i) subjective evaluation of pitch contours, ii) evaluation of pitch moments, iii) analysis of variation in mean pitch employing Student t tests, iv) analysis of variation in pitch variance employing F-tests, v) analysis of pitch distribution with respect to Gaussian and other speech styles employing Kolmogorov-Smirnov one and two sample tests.

Pitch Contours: Pitch contours from the Speaking Styles, Single Tracking Task/Lombard Effect domains were evaluated to determine time variation characteristics of pitch.

Pitch Moments: An evaluation of pitch moments was performed to uncover variation in: mean, variance, standard deviation, average deviation, skewness, and kurtosis. Results indicate that speech spoken under loud, anger, and question styles produced significantly larger mean pitch values when compared to neutral. Variance for all three styles were also significantly higher than neutral. Seperate evaluations confirmed pitch moment analysis.

Pitch Mean – Student t tests: *Student's t* tests were used to measure the significance of a differences in mean pitch across speaking styles. The significance of pairwise comparisons suggest that shifts in mean pitch appear to be good stress indicators over a wide variety of conditions. Secondary evaluations were performed to confirm statistical analysis. Of the 55 pairwise t tests performed, fifty-one resulted similar levels of significance. Those which differed involved analysis from single tracking task styles; thus indicating mean pitch to be an unreliable indicator for this type of task. Strong agreement across all other styles support the reliability of mean pitch as a stress indicator.

Pitch Variance – F-tests: *F-test* statistics measured the significance of difference of pitch variance across simulated stress domains. Significance from pairwise evaluations show pitch variance to be a good differentiating stress parameter. To confirm the analysis, secondary evaluations were performed. Of the 55 pairwise F-tests, 49 resulted in similar levels of significance. Those which differed involved analysis from slow or fast speaking styles, thus indicating pitch variance to be an unreliable stress indicator for speaking rate. Strong agreement across other styles supports the reliability of pitch variance as a stress indicator.

Pitch Distribution – Kolmogorov-Smirnov tests: Evaluation of pitch distribution was also considered. In the first evaluation, a *Kolmogorov-Smirnov one-sample* statistic was used to compare sample pitch distributions with Gaussian distributions with equivalent mean and variance. This was used to verify analysis employing parametric statistical tests.

Two-sample *Kolmogorov-Smirnov* tests were used to compare empirical pitch distribution functions across all simulated stress speaking styles. Significance results indicate pitch distribution to be a fair discriminating factor in differentiating speech under stress. Of the 55 tests, 37 resulted in paired distributions which were significantly different. Pitch distributions significantly different from most speaking styles include question, clear, Lombard, slow, and fast styles. Subsequent evaluations suggest pitch distribution to be a marginally reliable indicator of stress. The only departure were inconsistent results for moderate and high workload task conditions.

4.2 Glottal Source Characteristics

Aspects of each laryngeal pulse such as duration, instant of glottal closure, or pulse shape play important roles in a talker's ability to vary source characteristics. Analysis of spectral characteristics of the glottal flow waveform were considered in the analysis of speech under stress. Analysis of temporal variations are currently being investigated based on glottal inverse filtering [1].

An analysis of glottal source spectral characteristics was performed for each of the eleven simulated stress conditions. Distinguishing features of glottal flow spectra include spectral slope and amplitude. Linear regression analysis was performed for each estimated glottal source spectrum to obtain spectral tilt and energy distribution variation. A chi square $\chi^2(a,b)$ measure was used to determine the goodness-of-fit for the linear model. Results from regression analysis show slow, fast, soft, loud, angry, and Lombard conditions possess spectral tilt which was significantly different from neutral. Increased spectral energy and slope indicate that under certain stress conditions (loud, angry, and Lombard), glottal pulses will have irregular shapes, possessing sharp rise times and sharp corners. This is presumed to be the result of increased subglottal pressure, Bernoulli suction, and vocal-fold tension.

5 Conclusions

An evaluation of how speech characteristics change under varying levels of stress has been addressed with specific application to improving automatic isolated-word speech recognition. Results presented here focused on characteristics of pitch and glottal source spectrum. Further discussion in each of the five speech analysis domains can be found in [3]. Goodness-of-fit statistical tests were used to determine significance of parameter variation. Under certain conditions, parameters from various pitch and glottal source characteristics have been shown to be reliable stress indicators. These findings suggest that such parameter information can be used reliably to aid in automatic isolated-word speech recognition in stressful environments[4].

References

[1] K.E. Cummings, M.A. Clements, J.H.L. Hansen, "Estimation and Comparison of the Glottal Source Waveform Across Stress Styles Using Glottal Inverse Filtering," *Proc. of Southeastcon- 89,* Columbia, South Carolina, April 1989.

[2] J.H.L. Hansen and M.A. Clements, " Evaluation of Speech under Stress and Emotional Conditions," *Proc. of the Acoustical Society of America,* 114th Meeting, Miami, Florida, Nov. 1987.

[3] J.H.L. Hansen, " Analysis and Compensation of Stressed and Noisy Speech With Application to Robust Automatic Recognition," Ph.D. Thesis, Georgia Institute of Technology, 396 pages, July 1988.

[4] J.H.L. Hansen, M.A. Clements, " Stress and Noise Compensation Algorithms for Robust Automatic Speech Recognition," submitted to *Proc. 1989 IEEE ICASSP,* Glasgow, Scotland, U.K., May 1989.

[5] E. Lombard, " Le Signe de l'Elevation de la Voix, " *Ann. Maladies Oreille, Larynx, Nez, Pharynx,* Vol. 37, pp. 101-119, 1911.