Voice and speech variation under physical stress

Kyung-Wha Kim

Forensic Phonetician
Forensic Science Division
Supreme Prosecutor’s Office
Korea

Awarding Institution: Seoul National University, Korea
Date of Award: August 2007

KEYWORDS PHYSICAL STRESS, VOICE VARIATION, SPEECH VARIATION, ACOUSTIC ANALYSIS, AUTOMATIC SPEAKER RECOGNITION

Stress occurs in various types, levels and forms. It affects speech and causes negative effects on the performance of speech perception and speaker-recognition systems. In forensic phonetics, speech produced under stressful conditions may cause inaccuracies in voice identification. This thesis has two main purposes: the first aim is to investigate how the voice and speech of a speaker vary under physical stress; the second one is to examine how physical stress affects a speaker-recognition system. We selected 22 subjects (11 males, 11 females) with no health problems to participate in two experiments. They were selected by age group. Their mean age was 31.4 years and the distribution was as follows: ten were less than 30 years, eight were between 30 and 39, four were 40 and older.
Experiment 1 involved acoustic and phonetic analyses of the voices under stress-free and stressed conditions. The study subjects were asked to read a text under three different conditions: before stress (pre), right after stress (stress), and 15 minutes after stress (post). To simulate stress, the study subjects were asked to run up and down a set of stairs prior to the reading task. We measured blood pressure, pulse and respiration rate in order to verify physiological arousal. A range of parameters were considered as part of the phonetic and acoustic analysis. We measured the duration of speech (including and excluding pauses), the duration of pauses, the ratio of utterance and pause durations, frequency and place of pauses (unfilled and filled pauses), mean fundamental frequency (f0), Articulation Rate (AR), jitter and number of speech errors across the speaking conditions. We observed that each subject experienced significant increases in blood pressure, pulse and respiration rate due to the induced stress. The duration of pauses, both within sentence (wp) and between sentences (bp), was significantly longer and the duration of utterances was significantly shorter in stressed conditions. However, no significant effect of stress could be discovered in relation to number of pauses. It seems that, in response to stress, speakers rest longer in syntactically fixed pauses rather than increasing the number of pauses they use. Mean f0 significantly increased while the subjects were under stress. The increase in mean AR was significant due to the decrease of utterance duration in the stressed conditions. There were no significant differences for jitter and speech errors across the three conditions.

In this study, the observed differences in the voice parameters under observation can be accounted for by the variation in the physiological parameter of respiration. When speakers experienced stress, their respiration rate increased as they needed more oxygen to inhale and exhale right after the stress task. As a result, the duration of pauses, both wp and bp, increased. However, there was no significant difference in the number of pauses while under stress. It appears that speakers tend to use pauses taking into consideration the syntax and intonation unit, and prefer to increase the duration of each pause rather than increase the number of pauses. This led us to the conclusion that the duration of pauses may be a more reliable indicator for physical stress than number of pauses. As a consequence of the rise in respiration rate there was an increase in subglottal pressure during speech, which resulted in an elevation in mean f0 during the voiced sections.

We also investigated the effect of physical stress on speakers’ performance in cognitive tasks. For this, speakers completed four arithmetic tasks (two additions, two subtractions) verbally and, following this, they were interviewed in the stress-free and stressed conditions, respectively. It became evident that
males and females responded differently to stress. For male speakers, the number of wrong answers increased in the stressed condition; female speakers, on the other hand, showed an increase of wrong answers in the post-stress condition.

In experiment 2, we examined the intra-speaker variation using an automatic speaker-recognition program. For each speaker, we measured and compared normalisation scores between stressed and stress-free speech. The results indicated that the score of comparing pre-stress speech and post-stress speech is significantly higher than that of pre-stress and stressed speech. From this, we can derive that the variation of voice caused by physical stress affected the speaker-recognition results negatively.

We also compared stress-free and stressed speech between speakers. When a person’s voice changes considerably due to stress, it may cause an error in the automatic speaker-recognition program. This could result in two different voices having similar values for some acoustic parameters and, as a consequence, intra-speaker variability (comparison of an individual’s pre-stress and stressed speech) could be higher than inter-speaker variability (comparison of speech between two different speakers). In such cases, we have observed that the variation of mean $f_0$, max $f_0$, $f_0$ range, mean AR and duration of utterance was higher than the average value, and the performance of the speaker-recognition program is less precise. In addition, the possibility of errors was higher for female voices. This was due to the fact that the average normalisation scores for female voices were higher than those for male voices. Future research into the robustness of automatic speaker-recognition systems to differences in physical and psychological states of speakers, as is commonly found in forensic phonetic casework, is encouraged. The use of automatic speaker-recognition systems in combination with the traditional auditory-phonetic/acoustic approach could be beneficial to speaker-comparison work.