Analysis of the effects of physical task stress on the speech signal

Keith W. Godin and John H. L. Hansen[a)]
Center for Robust Speech Systems (CRSS), The University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080

(Received 10 September 2010; revised 11 September 2011; accepted 13 September 2011)

Physical task stress is known to affect the fundamental frequency and other measurements of the speech signal. A corpus of physical task stress speech is analyzed using a spectrum F-ratio and frame score distribution divergences. The measurements differ between phone classes, and are greater for vowels and nasals than for plosives and fricatives. In further analysis, frame score distribution divergences are used to measure the spectral dissimilarity between neutral and physical task stress speech. Frame scores are the log likelihood ratios between Gaussian mixture models (GMMs) of physical task stress and of neutral speech. Mel-frequency cepstral coefficients are used as the acoustic feature inputs to the GMMs. A Laplacian distribution is fitted to the frame scores for each of ten phone classes, and the symmetric Kullback–Leibler divergence is employed to measure the change in distribution from neutral to physical task stress. The results suggest that the spectral dissimilarity is greatest for the second level of a four level exertion measurement, and that spectral dissimilarity is greater for nasal phones than for plosives and fricatives. Further, the results suggest that different phone classes are affected differently by physical task stress. [DOI: 10.1121/1.3647301]

PACS number(s): 43.70.Fq, 43.72.Ar [AL] Pages: 3992–3998

I. INTRODUCTION

Exercise affects speakers, and the literature documents some of the resulting changes in the acoustic signal (Johannes *et al.*, 2007; Godin and Hansen, 2008). Exercise impacts the performance of speech systems (Entwistle, 2003), and it is known that speech, in turn, affects exercise (Meckel *et al.*, 2002). The full effects of exercise on the acoustic speech signal are not documented. This study presents a measurements analysis and an experiment to explore the effects of exercise on the speech spectrum, and to compare the relative effects of exercise on the phone classes of American English.

For consistency with related terminology, exercise is known in this study as physical task stress. Classification of the speaker related factors that affect the acoustic speech signal is nontrivial and motivated by varying contextual concerns (Murray *et al.*, 1996); thus, terminology has varied. In the past, speech under stress denoted any source of variability in the speech signal (Hansen, 1988). Several types of speech variability were considered in that study, including anger, shouting, Lombard effect, and workload stress. Solutions for speech recognition systems were explored in that study, and later studies also considered stress classification systems (Cairns and Hansen, 1994; Womack and Hansen, 1996; Bou-Ghazale and Hansen, 2000). Contemporary study of acoustic variability of speech focuses on one type of speech variability at a time, and groups emotions into one category, and situational speaker stressors into another, including cognitive task stress (Lindstrom *et al.*, 2008), Lombard effect (Boril and Hansen, 2010), fatigue (Vogel *et al.*, 2010), and physical task stress.

Godin and Hansen (2008) showed that fundamental frequency, and the percent of frames voiced in an utterance, are affected by physical task stress, but that the standard deviation of fundamental frequency in an utterance is not affected. Johannes *et al.* (2007) demonstrated that while heart rate and blood pressure increased linearly with increasing exertion level, fundamental frequency exhibits a nonlinear increase, characterized by multiple plateaus and abrupt transition functions between them. In experiments with a physiological microphone, a neck mounted contact transducer for signals up to 2 kHz, Patil and Hansen (2010) showed that such sensors afford both better speaker verification performance and better stress detection performance under physical task stress than typical acoustic close-talking microphones. Entwistle (2005) demonstrated the extent of the impact of physical task stress on the performance of speech recognition systems.

Meckel *et al.* (2002) investigated the effect of speech production on physiological variables, showing that, when under physical task stress, the addition of a speaking task results in increases in several physiological variables associated with exertion. Olson and Strohl (1987) accepted that nasal resistance decreases during exercise and investigated four possible causes, determining that none were the cause of the observed decreases. The effects of decreased airway resistance due to the use of helium-rich air mixtures is well known to undersea divers, raising formant center frequencies and voice pitch (Morrow, 1971), but the effects of decreased nasal resistance on the acoustic speech wave have not been investigated.

Rotstein *et al.* (2004) concluded that the perceived speech production difficulty—a nonacoustic measure—increases

[a)]Author to whom correspondence should be addressed. Electronic mail: john.hansen@utdallas.edu

 0001-4966/2011/130(6)/3992/7/$30.00

linearly with increasing exercise intensity as measured by three intensity metrics. Rotstein *et al.* (2004) also concluded that the "talk test"—a speech production-based measure of exercise intensity—is a "questionable substitute for the anaerobic threshold or [heart rate] for prescribing individual training exercise intensity." Based on this prior work, it is observed that while a given speech difficulty level may be associated with a given level of exertion, measured changes in speech production processes, or observed changes in the acoustic speech wave, may not be associated with any particular level of exertion across speakers.

Only Johannes *et al.* (2007) and Godin and Hansen (2008) analyzed the acoustic waveform of speech under physical task stress. The speech signal was treated as a whole; the impact of physical task stress on particular phone classes was not considered by either study. Section III of this study presents F-ratio measurements of the spectrum that show that different phone classes are affected differently by physical task stress, and that more sonorous phone classes, such as the vowels, appear to be more affected. Based on these observations, we hypothesize that a speech spectrum dissimilarity measure will be greater for the sonorant phone classes than for the nonsonorant phone classes. Based on the results of Meckel *et al.* (2002), Rotstein *et al.* (2004), and Johannes *et al.* (2007), we hypothesize that a speech spectrum dissimilarity measure will be greater for greater exertion levels. Based on the observation of Johannes *et al.* (2007) that fundamental frequency increases were characterized by plateaus separated by abrupt increases, four groups of exertion levels will be formed for the experiment. The experiment described in Sec. IV tests both of these hypotheses.

Next, Sec. II discusses the speech corpus that was used in this study. Section III presents measurements on the speech spectrum that motivate our hypothesis that speech spectrum dissimilarity will be greater for sonorant phone classes than for nonsonorant phone classes. Section IV presents an experiment to test both hypotheses in this study. Finally, Sec. V discusses the results.

II. CORPUS: UT-SCOPE

The UT-Scope corpus encompasses speech under Lombard effect, physical task stress, cognitive task stress, vocal effort, and whisper. It was collected by the Center for Robust Speech Systems (CRSS) at the University of Texas at Dallas. The physical task stress portion includes speech from 77 speakers, 51 of whom are native speakers of American English, 42 female and 9 male speakers (Ikeno *et al.*, 2007). For each speech type, speech was collected using 35 sentence prompts, prompted through headphones, and a 3 min spontaneous speech segment involving a conversation between the experimenter and the subject. The prompted segments of the recordings of the native speakers have full sentence and word level segmentations available, with phone-level segmentations available for 39 of the 42 female native speakers. The phone-level segmentations were generated using forced alignment.

The experiments in this study employed the prompted neutral and physical task stress segments of the 36 female native speakers for which both phone segmentations and heart rate data are available. Relevant aspects of the portion of UT-Scope used in this study are described in Table I. All speakers in UT-Scope were recorded with three microphones. This study uses data recorded from a Shure Beta 53 head-worn close-talking microphone. Heart rate was recorded using a Polar S520 heart rate watch with chest-worn sensor. Audio was recorded at 44.1 kHz to a Fostex D824 digital recorder, and downsampled to 16 kHz for this study. The speech was collected in an American Speech-Language-Hearing Association certified single-walled soundbooth. Physical task stress was induced using a Stamina Conversion II Elliptical/Stepper in the elliptical mode. Subjects were instructed to maintain an approximate 10 mph pace on the machine.

Some of the speakers in the corpus were more physically fit than others. Heart rate data (sampled at 15 s intervals) and the age of each speaker are used to form an estimate of the exertion level of each speaker. The formulation of the estimate incorporates the overall fitness level of the speaker by including the resting heart rate. Though it will tend to overestimate the resting heart rate, an estimate of the resting heart rate of each speaker is obtained by averaging the speaker's heart rate during the neutral segment. The exertion level is estimated using the Karvonen method (Davis and Convertino, 1975), which measures exertion along a percentage scale, where the bottom of the scale is the resting heart rate, and the top is an estimated maximum heart rate for that subject:

$$\mathrm{HR} = (\mathrm{MHR} - \mathrm{RHR})l + \mathrm{RHR}, \tag{1}$$

where HR is the current heart rate, RHR is the resting heart rate, l is the exertion level (ranging from 0.0 to 1.0), and MHR is the person's maximum heart rate (in beats/min), estimated according to (Tanaka *et al.*, 2001)

$$\mathrm{MHR} = 208.9 - 0.7A, \tag{2}$$

where A is the age of the person in years. Figure 1 shows the exertion level estimates for each speaker in the corpus. Figure 1 shows that the task employed to induce the stress is of medium difficulty for most speakers. The outlier in the graph at age 19 demonstrates the limitations of estimating maximum heart rate using age (Tanaka *et al.*, 2001), which result from formulating the estimation method using averages across many speakers. The average exertion level for the

TABLE I. Aspects of the subset of UT-Scope used in this study.

Gender of speakers	Female
Number of speakers	36
Average age (years)	23.6
Age range	18–45
Sentences/task	35
Tasks	Neutral, physical exertion
Native language	American English
Microphone	Close-talking
Speech style	Prompted
Av. exertion level	45.7%
Sampling rate	16 kH

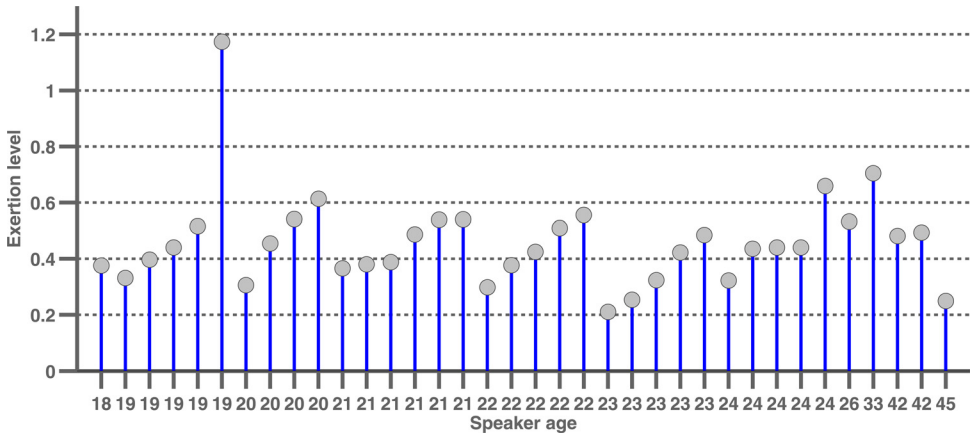

FIG. 1. (Color online) Exertion level for each speaker.

speakers in this corpus is 0.457, or 45.7%. Having established the utility of UT-Scope for the study of physical task stress speech, the next section focuses on an analysis of the changes in speech spectral structure.

III. F-RATIO MEASUREMENTS AND ANALYSIS

The following measurements were intended to provide insight into the differences between the effects of physical task stress on phone classes. For this purpose, F-ratio measurements were taken on the short time speech spectrum in order to examine qualitatively the effect of physical task stress on the mean of each spectral bin. Measurements were taken across all of the speech, as well as within subsets of the speech corresponding to ten classes of phones.

In the speech literature, Fisher's F-ratio has been applied similarly to compare general intra-speaker variability to inter-speaker variability (Lu and Dang, 2008), and has been applied to the measurement of the discriminative capability of features for speaker identification (Campbell, 1997). The F-ratio measurements were taken for each speaker, and then the measurements were averaged across speakers. An F-ratio measurement on the short time speech spectrum of a speaker was accomplished by forming the ratio of the variance between phone classes of each frequency bin of a discrete Fourier transform analysis to the average variance of that bin within the phone classes. For speaker j, spectrum bin i, the F-ratio is formulated in this study as

$$F_{i,j} = \frac{\frac{1}{M}\sum_{k=1}^{M}\left(u_{i,j,k} - u_i\right)^2}{\frac{1}{M}\sum_{k=1}^{M}\frac{1}{T_{j,k}}\sum_{l=1}^{T_{j,k}}\left(f_{i,j,k,l} - u_{i,j,k}\right)^2}, \tag{3}$$

where $f_{i,j,k,l}$ is the value of spectrum bin i for speaker j, category k, frame l, $T_{j,k}$ is the total number of frames from speaker j, category k, M is the number of categories (two, here), $u_{i,j,k}$ is the mean value of spectral bin i across all frames in category k of speaker j, and $u_{i,j}$ is the mean of bin i across all frames of speaker j. All analysis windows were 25 ms long, with a 10 ms window shift.

The F-ratio scales observed changes in the mean of each spectral bin by the average variance within each category. The F-ratio is sensitive to changes in the mean of a parameter; larger observed differences in means relative to within-category variance is associated with larger F-ratio measurements. The distributions of each spectral bin are best assumed multi-modal in this study, given, for example, the varying magnitudes of speech, and that each phone class contains multiple phones. This implies that the statistical significance of the F-ratio here cannot be measured by applying an F-distribution, because that would assume the underlying parameter is Gaussian distributed. Instead, the F-ratio provides a qualitative view of the differing effects physical task stress has across phone classes, as compared with neutral speech. These results motivate the quantitative analysis of the second experiment, discussed in Sec. IV.

The result of the F-ratio measurement taken across all phones is shown in Fig. 2. From Fig. 2, it can be observed that the effects of physical task stress are not concentrated in specific areas of the spectrum, though greater effects may be observed on very low and very high frequencies. The effects observed at lower frequencies may be related to fundamental frequency changes.

The spectrum F-ratio of the individual phone classes reveals that physical task stress affects different phone classes differently, and that all phone classes are affected. Ten phone classes were studied: High vowels, low vowels,

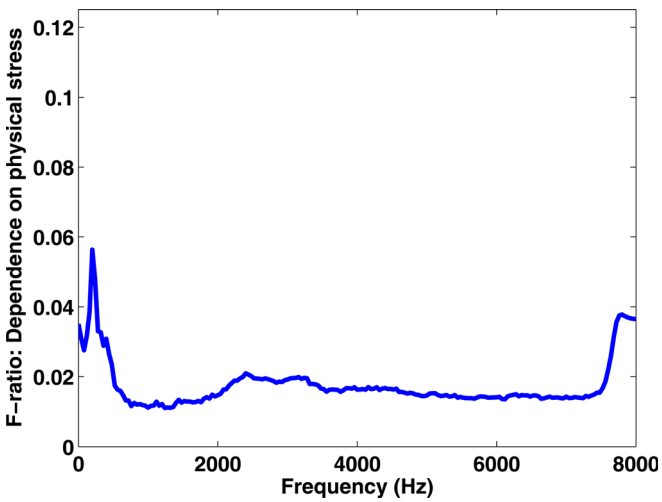

FIG. 2. (Color online) F-ratio showing change in mean of each spectrum bin due to physical task stress.

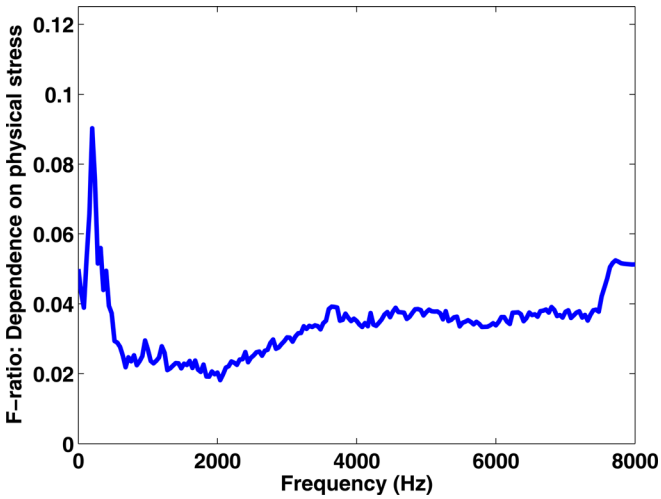

FIG. 3. (Color online) F-ratio showing change in mean of each frequency bin of high vowels due to physical task stress.

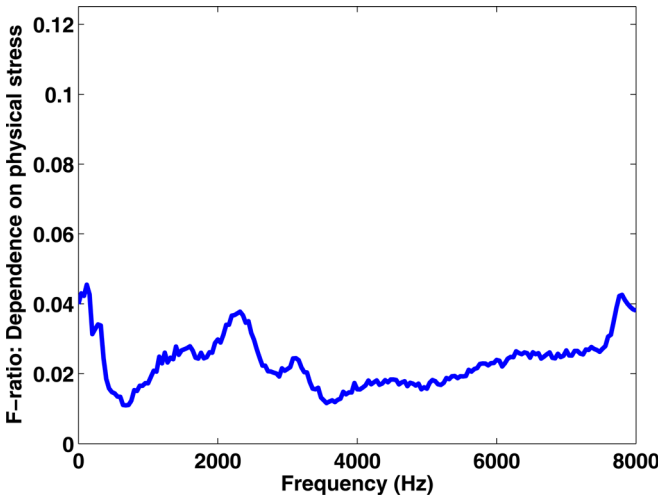

FIG. 5. (Color online) F-ratio showing change in mean of each frequency bin of fricatives due to physical task stress.

stops, fricatives, nasals, liquids, glides, diphthongs, and affricates. A representative set of five is presented: Figs. 3 and 4 show the average F-ratio for high vowels and low vowels, respectively, Figs. 5 and 6 show the average F-ratio for fricatives and stops, and Fig. 7 shows the average F-ratio for nasal phones. While the spectrum F-ratio of the stop plosives is similar to the overall F-ratio, appearing to vary little across the spectrum, the high vowels, low vowels, and nasals show not only greater overall change, but also a sharp peak at lower frequencies. Generally, the F-ratio is greater for voiced phones than unvoiced phones, and greater for more sonorous phones than less sonorous phones. Concluding whether these observed differences between the effects of physical task stress on different phone classes are statistically significant is a purpose of the experiment presented in Sec. IV.

IV. PHONE CLASS FRAME SCORE DISTRIBUTION ANALYSIS

The following experiment tested two hypotheses. It was hypothesized that speech spectrum dissimilarity between neutral and physical task stress speech will be greater for greater exertion levels, and that speech spectrum dissimilarity between neutral and physical task stress speech will be greater for sonorant phone classes than for nonsonorant phone classes. Preliminary results for this experiment were previously discussed in Godin (2009). In this experiment, the speech was first preprocessed by computing Mel-frequency cepstral coefficients (MFCCs) (Davis and Mermelstein, 1980), a frame-based acoustic feature, to reduce the data dimensionality, thus facilitating statistical modeling of the acoustic waveform. Next, a pair of statistical models was employed to estimate the difference in log posterior probabilities, a "score," that represents an estimate of whether the frame exhibits particular characteristics of physical task stress speech, or of neutral speech. Frames more closely associated with physical task stress have higher scores than frames more closely associated with neutral speech.

It is from these frame scores that the measurement of the effects of physical task stress on specific phone classes was made. Scores from neutral speech for one phone class

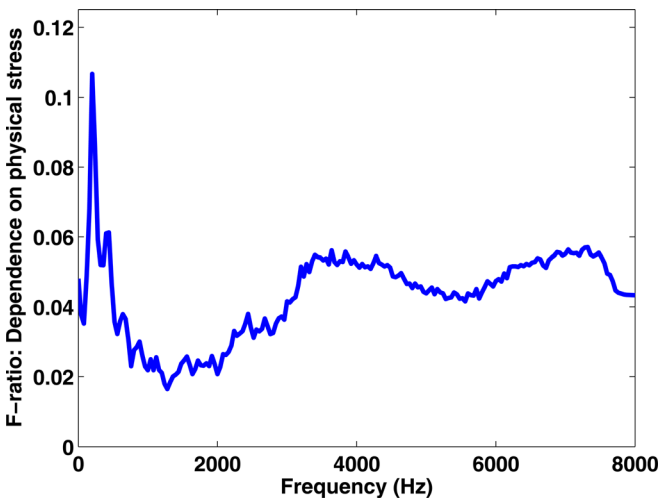

FIG. 4. (Color online) F-ratio showing change in mean of each frequency bin of low vowels due to physical task stress.

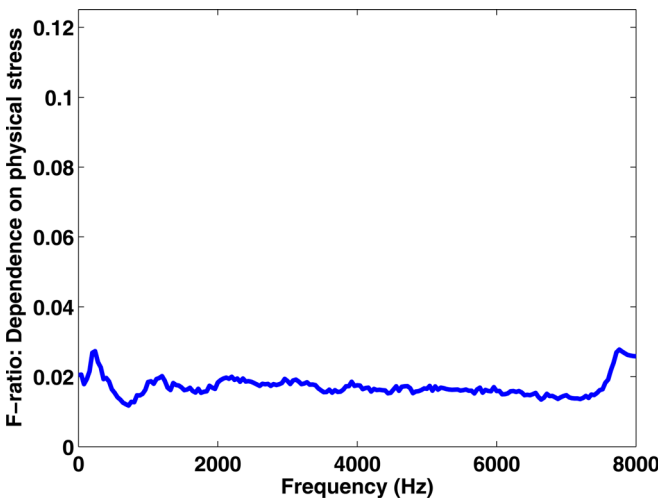

FIG. 6. (Color online) F-ratio showing change in mean of each frequency bin of stop plosives due to physical task stress.

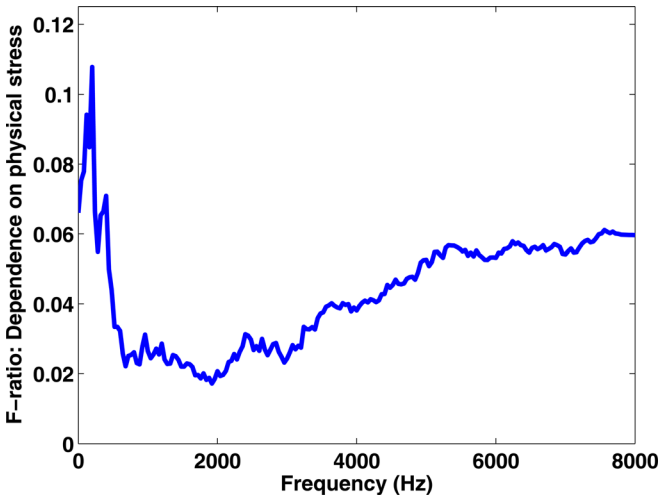

FIG. 7. (Color online) F-ratio showing change in mean of each frequency bin of nasals due to physical task stress.

for one speaker were compared with scores from physical task stress speech for that same phone class for that same speaker by estimating and comparing statistical models for each. An analysis of variance (ANOVA) was employed to determine the significance of these score comparisons. Some details of the experimental design follow.

A. Estimating frame scores

Gaussian mixture models (GMMs) (Reynolds and Rose, 1995) were employed to model the distribution of acoustic features for neutral and for physical task stress speech. The GMMs were trained using the Hidden Markov Model Toolkit [(HTK), Young *et al.* (2006)]. From these, posterior probabilities that a given frame is produced by the physical task stress model and the neutral speech model were computed and compared, resulting in the frame score.

A separate pair of GMMs was estimated for each speaker in the experiment, in the sense that the GMM parameters are estimated by pooling all other speakers' data as training data for the model pair. The number of mixtures to use for the GMMs was determined by treating the GMM pairs as stress/neutral classification systems, applying the classification system to the utterances of the left-out speaker, and increasing the number of mixtures by a factor of 2 until the classification error rate stopped decreasing. Details of the resulting GMM set are shown in Table II.

B. Comparing phone class score distributions

Measuring the effects of physical task stress on a phone class was achieved by comparing the estimated distribution of frame scores for that phone class in neutral with an estimated distribution of frame scores for that phone class in physical task stress. This comparison of distributions was performed using the symmetric Kullback–Leibler (KL) divergence. The KL divergence was used because it is a general, straightforwardly applied measure related to changes in the respective probability distributions. All phone class frame scores are assumed (justified below) to be Laplacian distributed, and thus the KL divergence was computed analytically using the parameters of a fitted Laplacian distribution.

The KL divergence compares two probability distributions $p(x)$ and $q(x)$,

$$D_{\mathrm{KL}}(P\|Q) = \int_{-\infty}^{\infty} p(x)\log\frac{p(x)}{q(x)}dx. \tag{4}$$

In this form, the KL divergence is not symmetric. The KL divergence is symmetrized by

$$D_{\mathrm{KL}}(P\,Q) = D_{\mathrm{KL}}(P\|Q) + D_{\mathrm{KL}}(Q\|P). \tag{5}$$

The Laplacian distribution is

$$p(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right). \tag{6}$$

The comparison equation is found by substituting Eq. (6) into Eq. (4), then substituting the result into Eq. (5). A complete derivation may be found in Godin (2009).

The result is

$$D_{\mathrm{KL}}(N,S) = b_n + b_s - b_n\,\exp\left(\frac{-|\mu_n-\mu_s|}{b_n}\right) - b_s\exp\left(\frac{-|\mu_n-\mu_s|}{b_s}\right), \tag{7}$$

where parameters b_n and μ_n are the parameters of the Laplacian distribution for the neutral frame scores, and parameters b_s and μ_s are the parameters of the Laplacian distribution for the physical task stress frame scores. The parameter μ is estimated as the median of the available data samples, and the maximum likelihood estimate of b is

$$\hat{b} = \frac{1}{M}\sum_{n}|s[n]-\mu|, \tag{8}$$

where M is the number of frames for that phone class, and n iterates through those frames.

Table III shows the number of speakers for whom Kolmogorov–Smirnov (KS) tests supported the assumption that the frame scores for that phone class are Laplacian distributed. For most phone classes, frame scores from a

TABLE II. Aspects of the classification system used in this study.

Parameter	Value used in experiment
Features used	MFCCs
Number of mixtures	256
Number of cepstral coefficients	15
Delta coefficients	√
Double-delta coefficients	√
Training software	HTK
Training speakers	41
Test speakers	1
Testing style	Round-robin
Equal error rate	15%
Global threshold	–0.1670

TABLE III. Results of statistical tests to determine whether frame scores within each phone class for each speaker are Laplacian distributed at the 99% confidence level. Thirty-nine total speakers.

Phone class	Number of speakers' scores Laplacian distributed
Neutral—low non-R vowels	33
Phy—low non-R vowels	34
Neutral—high non-R vowels	22
Phy—high non-R vowels	25
Neutral—laterals	38
Phy—laterals	38
Neutral—stop plosives	33
Phy—stop plosives	35
Neutral—diphthongs	35
Phy—diphthongs	33
Neutral—R vowels	36
Phy—R vowels	36
Neutral—fricatives	32
Phy—fricatives	29
Neutral—glides	35
Phy—glides	35
Neutral—nasals	31
Phy—nasals	34
Neutral—combo consonants	35
Phy—combo consonants	35

majority of the speakers are consistent with a Laplacian distribution. This justifies the application of the assumption that the frame scores for all phone classes for all speakers are Laplacian distributed. KS tests for the fit of a Gaussian distribution to the frame scores found that assuming a Gaussian distribution for frame scores was a reasonable assumption for less than 10% of the speakers, for most phone classes.

C. Results

Figures 8 and 9 show the results of the experiment. A two-way ANOVA, shown in Table IV, was performed on the results, with phone class and exertion level as main effects. Exertion level was grouped into four levels, rather than modeled as a linear continuous parameter, consistent with the hypothesis that the relationship between spectrum dissimilarity and exertion level is nonlinear. Thresholds for the exertion level categories were set to balance the number of speakers in each group. The confidence level for the ANOVA was set at 95%.

Table IV shows that both phone class and exertion level are both significant main effects. The interaction between phone class and exertion level was not significant. *Post hoc*, pairwise comparisons were computed using the Tukey–Kramer method, and the resulting confidence intervals for comparison are shown in both Figs. 8 and 9. The difference between the effects on nasals versus the effects on both fricatives and plosives is statistically significant. Pairwise comparison of exertion levels showed that the frame score distribution changes for the second level are significantly lower than for the other three levels.

V. DISCUSSION

Based on the results of the second experiment, we cannot accept our hypothesis that spectral dissimilarity is greater for the sonorant phone classes. Instead, we can conclude that the average spectra of nasal phones are more affected by physical task stress than the average spectra of plosives and fricatives. One possible mechanism for this may be the decrease in nasal resistance associated with physical task stress as observed by Olson and Strohl (1987), which may result in a change in center frequencies of nasal resonances or anti-resonances. A future study could explore the hypothesis that physical task stress results in changes in the resonances observed in the acoustic wave of nasal phones.

We also cannot accept our hypothesis that spectral dissimilarity is correlated with greater exertion levels. However, our results suggest a relationship between spectral dissimilarity and exertion. One possible mechanism for this difference may be differing choices of breathing strategy among different exertion levels. A future investigation could

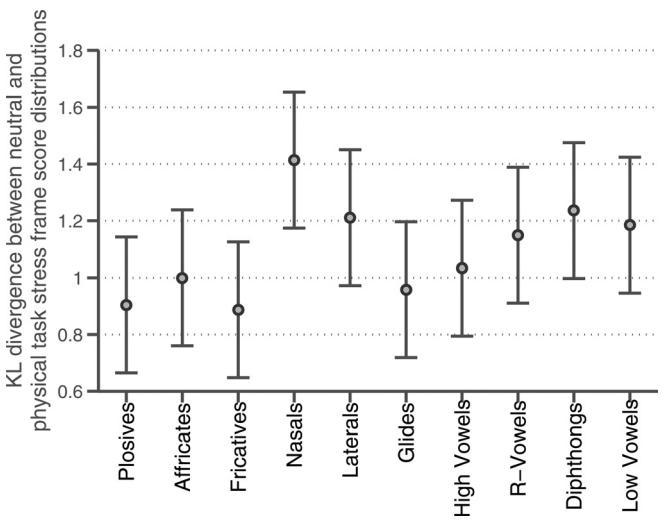

FIG. 8. Average effect of physical task stress on change in frame score distribution, grouped by phone class and ordered by sonority. Bars show pairwise comparison intervals.

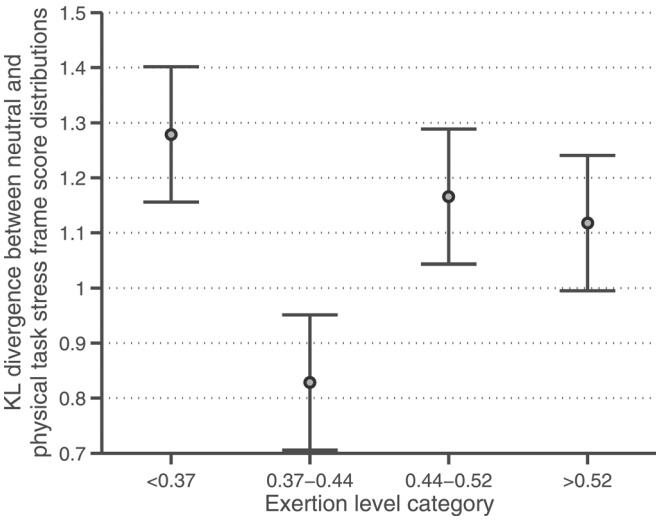

FIG. 9. Average effect of physical task stress on change in frame score distribution, grouped by exertion level. Bars show pairwise comparison intervals.

TABLE IV. Results of two-way ANOVA with phone class and exertion as main effects.

	SumSq.	D.F.	MeanSq.	*F*	*p*
Phone class	8.35	9	0.928	2.39	0.0124
Exertion	4.021	2	2.01	5.17	0.0062
Phone class* exertion	3.51	18	0.195	0.50	0.9562
Error	120.204	310	0.388		
Total	136.47	339			

explore the hypothesis that when faced with a task of medium to low difficulty, speakers use a different breathing strategy that involves pushing their breaths out to the edges of a sentence, and thus maintaining less altered speech production.

The statistically insignificant differences observed in Fig. 8 suggest further investigation into the possibility that the diphthongs and low vowels are more affected by physical task stress than the plosives, affricates, and fricatives, by exploring in particular the effects of physical task stress on the voice production system. Nonacoustic measures of the behavior of the voice production system could assist in further exploring these effects in future work. An electroglottograph, for example, could isolate the effects of physical task stress on the fundamental frequency from acoustic measurement errors related to increased noise in the vocal tract; it could also be employed to compare speech under physical task stress with pressed voicing, or be used to perform inverse filtering to more precisely measure formant center frequencies.

ACKNOWLEDGMENTS

This project was supported by AFRL through a subcontract to RADC Inc. under FA8750-09-C-0067, and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J.H.L.H.

Boril, H., and Hansen, J. H. L. (**2010**). "Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments," IEEE Trans. Audio, Speech, Lang. Process. **18**, 1379–1393.

Bou-Ghazale, S. E., and Hansen, J. H. L. (**2000**). "A comparative study of traditional and newly proposed features for recognition of speech under stress," IEEE Trans. Speech Audio Process. **8**, 429–442.

Cairns, D. A., and Hansen, J. H. L. (**1994**). "Nonlinear analysis and classification of speech under stressed conditions," J. Acoust. Soc. Am. **96**, 3392–3400.

Campbell, J. P. (**1997**). "Speaker recognition: A tutorial," Proc. IEEE **85**, 1437–1462.

Davis, J. A., and Convertino, V. A. (**1975**). "A comparison of heart rate methods for predicting endurance training intensity," Med. Sci. Sports **7**, 295–298.

Davis, S. P., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**, 357–366.

Entwistle, M. S. (**2003**). "The performance of automated speech recognition systems under adverse conditions of human exertion," Int. J. Hum.-Comput. Interact. **16**, 127–140.

Entwistle, M. S. (**2005**). "Training methods and enrollment techniques to improve the performance of automated speech recognition systems under conditions of human exertion," Ph.D. thesis, University of South Dakota, Vermillion, SD.

Godin, K. W. (**2009**). "Classification based analysis of speech under physical task stress," Master's thesis, University of Texas at Dallas, Richardson, TX.

Godin, K. W., and Hansen, J. H. L. (**2008**). "Analysis and perception of speech under physical task stress," in *INTERSPEECH 2008*, Brisbane, Australia, pp. 1674–1677.

Hansen, J. H. L. (**1988**). "Analysis and compensation of stressed and noisy speech with application to robust automatic recognition," Ph.D. thesis, Georgia Institute of Technology, Atlanta.

Ikeno, A., Varadarajan, V., Patil, S., and Hansen, J. H. L. (**2007**). "UT-Scope: Speech under Lombard effect and cognitive stress," in *IEEE Aerospace Conference 2007*, Big Sky, MT, pp. 1–7.

Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C. A., Thomas, J. L., Adler, A. B., and Gerzer, R. (**2007**). "Non-linear function model of voice pitch dependency on physical and mental load," Eur. J. Appl. Physiol. **101**, 267–276.

Lindstrom, A., Villing, J., Larsson, S., Seward, A., Aberg, N., and Holtelius, C. (**2008**). "The effect of cognitive load on disfluencies during in-vehicle spoken dialogue," in *INTERSPEECH 2008*, Brisbane, Australia, pp. 1196–1199.

Lu, X., and Dang, J. (**2008**). "An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification," Speech Commun. **50**, 312–322.

Meckel, Y., Rotstein, A., and Inbar, O. (**2002**). "The effects of speech production on physologic responses during submaximal exercise," Med. Sci. Sports Exercise **34**, 1337–1343.

Morrow, C. T. (**1971**). "Speech in deep-submergence atmospheres," J. Acoust. Soc. Am. **50**, 715–728.

Murray, I. R., Baber, C., and South, A. (**1996**). "Towards a definition and working model of stress and its effects on speech," Speech Commun. **20**, 3–12.

Olson, L. G., and Strohl, K. P. (**1987**). "The response of the nasal airway to exercise," Am. Rev. Respir. Dis. **135**, 356–359.

Patil, S. A., and Hansen, J. H. L. (**2010**). "The physiological microphone (PMIC): A competitive alternative for speaker assessment in stress detection and speaker verification," Speech Commun. **52**, 327–340.

Reynolds, D. A., and Rose, R. C. (**1995**). "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. Speech Audio Process. **3**, 72–83.

Rotstein, A., Meckel, Y., and Inbar, O. (**2004**). "Perceived speech difficulty during exercise and its relation to exercise intensity and physiological responses," Eur. J. Appl. Physiol. **92**, 431–436.

Tanaka, H., Monahan, K. D., and Seals, D. R. (**2001**). "Age-predicted maximal heart rate revisited," J. Am. Coll. Cardiol. **37**, 153–156.

Vogel, A. P., Fletcher, J., and Maruff, P. (**2010**). "Acoustic analysis of the effects of sustained wakefuleness on speech," J. Acoust. Soc. Am. **128**, 3747–3756.

Womack, B. D., and Hansen, J. H. L. (**1996**). "Improved speech recognition via speaker stress directed classification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, pp. 53–57.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2006**). "The HTK book (for HTK version 3.4)," pp. 300–302, http://htk.eng.cam.ac.uk (Last viewed: June 5, 2010).