Analysis of Cognitive Loaded Speech using Excitation Source Features

Sudarsana Reddy Kadiri[1], **P. Gangamohan** [2], **Suryakanth V Gangashetty** [3] **and B. Yegnanarayana** [4]
Speech and Vision Laboratory,
Language Technologies Research Center,
International Institute of Information Technology-Hyderabad, India.
[1]sudarsanareddy.kadiri@research.iiit.ac.in, [2]gangamohan.p@students.iiit.ac.in,
[3]svg@iiit.ac.in,[4]yegna@iiit.ac.in

Abstract

Cognitive loaded speech is produced when a speaker experiences the load imposed by a certain task on the cognitive system and it can be regarded as deviation from neutral speech. The objective of the present study is to explore the deviations in the excitation source features of cognitive loaded speech compared to neutral speech. The excitation source features considered in this study are, Instantaneous fundamental frequency (F_0), Strength of excitation (SoE), Energy of excitation (EoE) and Perceived loudness (η). For the extraction of these excitation source features, two signal processing techniques are used, namely Zero frequency filtering (ZFF) and Linear Prediction (LP) analysis. A comparative study of excitation source features across different cognitive loaded speech indicates that there are significant deviations in the excitation source features of cognitive loaded speech compared to neutral speech.

Keywords: Cognitive Loaded Speech, Excitation Source Features, Zero Frequency Filtering, Linear Prediction Analysis.

1 Introduction

The cognitive load on a person refers to the amount of mental demand imposed when performing a particular task (Lively, 1993; Yap, 2012). The central theme to cognitive load theory is based on the assumption that the human working memory is limited and hence, as a task becomes more difficult, the availability of working memory is reduced and hence cognitive load increases (Yap, 2012; Le et al., 2011). Recent studies on cognitive load indicates that, the performance of the user will degrade if the load is too high or too low (Post et al., 2013; Kirschner et al., 2009). However, it may be possible to adjust the workload of the person, if the cognitive load level can be estimated and, hence the productivity can be improved by designing adaptive Human-Computer interfaces.

Earlier studies have revealed that, the influence of cognitive load on speech reflects in disfluencies, articulation rate, content quality, number of syllables, silent pauses, filled pauses, average pause length, average pause frequency, and average response latency etc. (Yin et al., 2008a; Khawaja et al., 2009; Scherer et al., 2002). However, extraction of these high level features is an expensive task. In many of the recent studies (Chen et al., 2011; Yin et al., 2008a; Yin et al., 2008b; Yin et al., 2008c; Yap et al., 2011a; Yap et al., 2010b; Yap et al., 2010a; Le et al., 2010; Boril et al., 2010), analysis of cognitive loaded speech is performed using mostly spectral parameters such as Mel frequency cepstral coefficients (MFCCs), Linear prediction coefficients (LPCs), Perceptual linear prediction coefficients (PLPs), Formant frequencies and their bandwidths etc (Yap et al., 2011a; Yap et al., 2010b; Yap et al., 2010a; Le et al., 2010; Boril et al., 2010). The basic analysis of spectral features, as reported in (Yap et al., 2011a; Yap et al., 2010b; Yap et al., 2010a), is shown in Table 1.

The spectral features (such as MFCCs, formant frequencies etc.) along with the prosody features (such as pitch and intensity etc.) has shown a performance improvement of cognitive load classification (Yin et al., 2008a; Yin et al., 2008b; Le et al., 2011; Yap et al., 2009). Apart from pitch and formant frequencies, low level features characterizing the spectral energy distribution have also been found to be indicative of cognitive load (Le et al., 2011; Boril et al., 2010). It is also observed that, the increase in cognitive load is reflected by an increase in spectral energy spread and spectral centre of gravity and the gradient of energy decay (Le et al., 2011; Scherer et al., 2002). From the analyses of spectral distribution information for cognitive loaded speech, it was found that most of the amount of cognitive load specific information was in below the 1 kHz frequency band (Le et al., 2011; Yin et al., 2008a; Yin et al., 2008b; Yin et al., 2008c). The non-linear Teager energy operator was found to be effective for classifying cognitive loads (Fernandez and Picard, 2003).

In (Yap et al., 2009), the authors exploited the phase characteristics of the speech using group delay features and found that these features are complimentary to the MFCCs for cognitive load classification. The features such as perceptual linear prediction coefficients, spectral centre of gravity, spectral energy spread, and the vowel durations were also found to be effective for the development of cognitive load classification systems

Table 1: *Basic spectral analysis of cognitive loaded speech with respect to neutral (Low load) speech, (inc: Increase, dec: Decrease).*

	F_1 mean	F_2 mean	F_3 mean	F_4 mean
Medium Load	*inc*	*dec*	–	*inc*
High Load	*inc*	*dec*	–	*inc*

(Boril et al., 2010).

Among all the features that have been used for cognitive load classification, MFCCs were recognized as one of the most effective features when the shift delta coefficients (SDC) of the features are used (Yin et al., 2008a; Yin et al., 2008b; Yin et al., 2008c). From the assumption that MFCCs may not capture the characteristics of the spectral envelope completely, spectral energy distributions within mel filter subbands are exploited in each subband and found that spectral centroid features can captures the more information, and they appear to be complimentary to the MFCCs. Motivated by this, in (Le et al., 2011), it was shown that the fusing of spectral centroid features with the MFCCs and prosody features provides an improvement in cognitive load classification.

The exploration of excitation source features is not up to the extent as it is for spectral features of cognitive loaded speech, even though excitation source based features have been found to be useful in other areas of speech such as speaker recognition, expressive speech processing, speech synthesis etc. (Drugman et al., 2014; Gangamohan et al., 2013). In (Yap et al., 2011b; Yap et al., 2010c), authors made an attempt to see the effectiveness of glottal flow (excitation source) features for cognitive load analysis and have found that the glottal source contains information for cognitive load discrimination. The features used are open quotient, normalized amplitude quotient and speed quotient extracted using the Glottal inverse filtering (Yap et al., 2010c). However, the reliability of glottal flow features depends on the accuracy of the glottal flow estimation, which is a non-trivial process (Drugman et al., 2014; Alku, 2011; Yegnanarayana and Murty, 2009).

Hence, in this paper, we propose to use the excitation source features extracted directly from the speech signal and do not require the explicit estimation of glottal flow. The features considered are obtained using analysis at epoch locations of the excitation source. They are, Instantaneous fundamental frequency (F_0), Strength of excitation (SoE), Energy of excitation (EoE) and Perceived loudness (η) and they reflects the characteristics of the vocal fold vibration.

The organization of the paper is as follows. Section 2 gives the basis for the present study. In Section 3, details of the data and feature extraction procedure is given. The analysis of excitation source features and discussion on results of the experiments are given in Section 4. Finally, Section 5 gives a summary and scope for further study.

2 Basis for the Present Study

As the cognitive loaded speech is produced by the human speech production mechanism, it has to be analyzed like a speech signal based on both the excitation source and vocal tract system characteristics. Due to the extra effort required to produce the cognitive loaded speech, the primary effect is expected to lie on the source of excitation due to pressure from the lungs and the vibration of the vocal folds. One motivation for the present study is from the studies (Steeneken, 1999), where the physiological consequences of the mental workload are analyzed such as, increased respiration rate, irregular breathing, and increased muscle tension of the vocal cords and shown that the increases in cognitive load have been associated with increases in pitch (Scherer et al., 2002; Boril et al., 2010; Steeneken, 1999). The other motivation for the present study came from the recent studies (Yegnanarayana and Murty, 2009; Murty and Yegnanarayana, 2008; Yegnanarayana and Gangashetty, 2011; Seshadri and Yegnanarayana, 2009), where the excitation source features can be extracted directly from the speech signal unlike most of the studies uses the glottal inverse filtering which is a non-trivial process (Alku, 2011; Drugman et al., 2014; Murty and Yegnanarayana, 2008). Hence, in this study the changes in the excitation source features that reflect the characteristics of vocal fold vibration are examined for cognitive loaded speech.

3 Speech Data and Feature Extraction

3.1 Speech Databases

For the analysis of cognitive loaded speech using excitation source features, the Cognitive Load with Speech and EGG (CLSE) database is chosen (Schuller et al., 2014; Yap, 2012). It was recorded using a close-talk microphone sampled at 16 kHz from 26 native Australian English speakers (20 male and 6 female). The database also contains electroglottograph (EGG) signals which was recorded simultaneously with speech. The CLSE database contains tasks such as reading span task, Stroop test with time pressure and the Stroop test with dual task. These tasks contains the speech corresponding to the three cognitive load levels such as Low or neutral (L1), Medium (L2) and High (L3) cognitive levels. More details on the database can be found in (Schuller et al., 2014; Yap, 2012).

3.2 Feature Extraction

The features related to the excitation source component of speech signal are used in this study. The significant excitation of the vocal tract system takes place during the abrupt closure of the vocal folds and it is almost like impulse. Hence, the high SNR of speech is present around the epochs (GCIs). The importance of anchoring the analysis around the glottal closure (high SNR regions) for processing the speech signal has been extensively covered in the recent articles (Drugman et al., 2012; Yegnanarayana and Gangashetty, 2011; Gangamohan et al., 2013). The features investigated in the present study are Instantaneous F_0, SoE, EoE and η extracted around the GCIs of speech signal. For this purpose, we use two signal processing methods, one is, a recently proposed method, Zero Frequency Filtering (ZFF) (Yegnanarayana and Gangashetty, 2011) and another is LP analysis (Makhoul, 1975).

3.3 Zero Frequency Filtering (ZFF) Method

The idea behind this study was that, the effect of impulse-like excitation is reflected across all frequencies including zero frequency (0 Hz) of the speech signal. Hence, the method involves the passing the speech signal through a cascade of two ideal digital resonators located at 0 Hz and followed by trend removal using the average pitch period. The trend removed signal is called zero frequency filtered (ZFF) signal. The instants of negative-to-positive zero crossings (NPZCs) of the ZFF signal correspond to the instants of significant excitation, i.e., *epochs or Glottal Closure Instants (GCIs)* in voiced speech (Murty and Yegnanarayana, 2008; Yegnanarayana and Gangashetty, 2011).

3.3.1 Instantaneous Fundamental Frequency (F_0)

The instantaneous fundamental frequency (F_0) is extracted using the ZFF method (Yegnanarayana and Murty, 2009; Yegnanarayana and Gangashetty, 2011). The interval between successive epochs (GCIs) gives the value of instantaneous pitch period (T_0) and hence the instantaneous fundamental frequency is given by $F_0 = 1/T_0$.

3.3.2 Strength of Excitation (SoE)

The slope of the ZFF signal around the epochs corresponds to the strength of excitation (SoE), which is closely related to the strength of the impulse-like excitation at epoch (Murty et al., 2009).

3.3.3 Voiced/unvoiced Detection

ZFF method is also useful for detecting voiced and unvoiced regions. Since the significant contribution by the impulse-like excitation, ZFF signal energy is high in voiced regions. The strength of the epochs and the energy of the ZFF signal are used to determine the regions of voiced and unvoiced segments (Dhananjaya and Yegnanarayana, 2010). In the present study, only voiced regions are considered for the analysis.

3.4 Linear Prediction (LP) Analysis

LP analysis with proper LP order gives the excitation source (LP residual) component and vocal tract system component (LPCs) (Makhoul, 1975). In the present study, for the extraction of excitation source features such as Energy of excitation (EoE) and Perceived loudness (η), LP residual is used. For this, a 10th order LP analysis is performed on the signal sampled at 8 kHz.

3.4.1 Energy of Excitation (EoE)

The EoE feature is computed using the energy of the samples of the Hilbert envelope of LP residual over 2 ms around each GCI and it gives an indication of the vocal effort.

3.4.2 Loudness Measure (η)

The measure of the abruptness of glottal closure was proposed in (Seshadri and Yegnanarayana, 2009) as a feature of loudness in speech. The loudness measure (η) is defined as the ratio of the standard deviation and mean of the samples of the Hilbert envelope of LP residual around each epoch.

As an illustration, the instantaneous F_0, SoE, EoE and η contours corresponding to voiced regions of *Low Cognitive Load level (L1)* utterance are shown in Fig. 1.

4 Feature Analysis and Discussion

It is intuitive and also observed that F_0 of medium and high cognitive loads are high with respect to low cognitive load (neutral) level and it is in conformity with the studies (Steeneken, 1999; Le et al., 2011; Le et al., 2010). Even though the average F_0 values for medium and high cognitive loads are high with respect to neutral, it is observed that the medium load has slightly less average F_0 compared to high cognitive loaded speech. It is observed that, even though the SoE is increasing as the cognitive load is increasing, the variance of the high cognitive load is very low and it is more concentrated when compared with the medium load. The EoE is also found to be increasing as the cognitive load level is increasing. Even though the loudness measure is showing increasing behavior, it is observed that, it is more of speaker specific. The summary of the analysis of excitation source features of cognitive load level speech with respect to neutral (low load) is given in Table II and in Table III the statistical analysis of the excitation source features is given for reading span task. From the observations in Tables II and III, it is indicative that the excitation source features are showing discrimination for cognitive loaded speech.

The present study is speaker-specific, where reference (neutral) and test utterances (any of cognitive load level) of the same speaker are considered together. In order to capture the relations among the features, distributions of two features are considered at a time.

Six 2-dimensional (2-D) feature spaces are formed by the combination of two features at a time: $C1$:

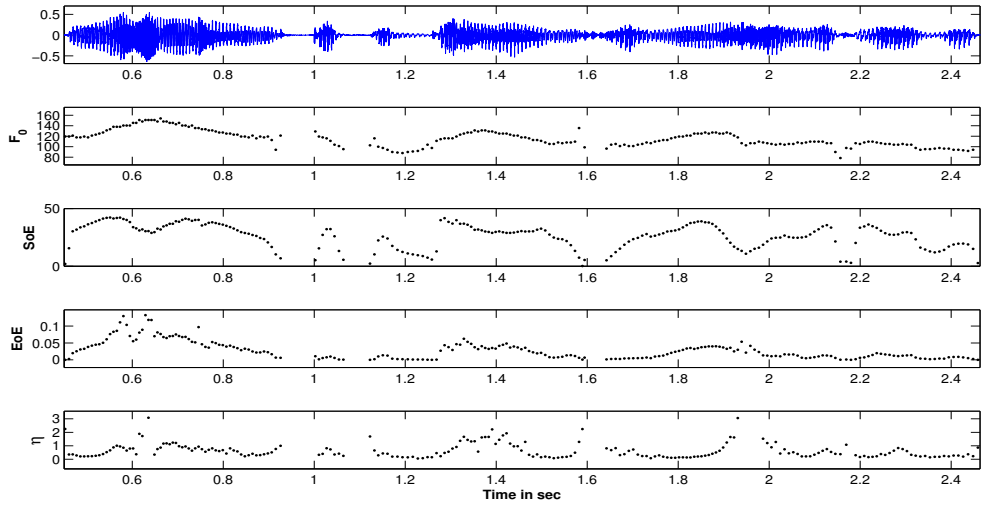

Figure 1: *A segment of Low Cognitive Load level (L1) utterance and its instantaneous F_0, SoE, EoE and η contours in voiced regions.*

Table 2: *Excitation source feature analysis of cognitive loaded speech with respect to neutral (Low load (L1)) speech.*

	F_0	SoE	EoE	η
Medium (L2)	High	High	High	High
High (L3)	Very High	Very High	Very High	High

$(F_0\ vs\ SoE)$, $C2 : (F_0\ vs\ EoE)$, $C3 : (F_0\ vs\ \eta)$, $C4 : (SoE\ vs\ EoE)$, $C5 : (SoE\ vs\ \eta)$ and $C6 : (EoE\ vs\ \eta)$, for each of a given set of reference (neutral) and test (any of cognitive load level) utterances. For an illustration, the 2-D feature spaces for a reference (neutral (L1)) and test (high cognitive load level (L3)) utterances are shown in Fig. 2.

In order to measure the divergence between the reference and test utterances, Kullback-Leibler (KL) distance between 2-D feature spaces are computed by modeling the feature space with Gaussian probability distribution function, which is represented by mean vector and covariance matrix. The measure of KL distance is given as

$$D = \frac{1}{2}(tr(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T\Sigma_1^{-1}(\mu_1 - \mu_0)) - \frac{1}{2}(k + ln(\frac{det\Sigma_0}{det\Sigma_1})) \quad (1)$$

where, D is the KL distance, k is the dimension of the distribution, Σ_0, Σ_1 are the covariance matrices of feature pair distributions of reference and test utterances and μ_0, μ_1 are the corresponding mean vectors, respectively.

Evaluations are performed on the three cognitive loaded speech tasks. Here, we considered 3 utterances for each cognitive load level. The average KL distance values for a set of reference (neutral) utterance and test (any of cognitive load level) utterances for each 2-D feature space are shown in Table 4. The KL distance values between the reference *neutral* utterance and the test *neutral* utterances are lower, when compared to the distance values between the reference *neutral* and the test (medium or high cognitive load level) utterances. This observation indicates that the speakers modify their vocal effort in producing the cognitive loaded speech.

It is also observed that the KL distance values depend on the feature space combination, speaker and the cognitive load level. For example, the KL distance values of all feature combinations in the case of high cognitive load (test) utterances are high most of the times, which can be seen from Tables 4,5 and 6. This indicates that high cognitive load level has more deviations from neutral. It is also observed that the absolute KL distance values are varies with respect to speaker. The KL distance values are lower when loudness measure is included in the feature combinations. This indicates that loudness is not changing too much in the production of cognitive loaded speech. The KL distance values are very low for all feature combinations when both the reference and test utterances are neutral, it indicates that, the neutral state is sustainable in terms of the excitation source. The variation in the KL distance values when the test utterance is cognitive loaded speech (L2 or L3) indicates that, there are unsustainable regions in cognitive loaded speech. Based on these observations, we can say that the analyzed excitation source features carry the production information of cognitive loaded speech. These observations indicate that feature corresponding to excitation source component are important

Table 3: *Statistics of excitation source parameters of cognitive loaded speech for reading span task (where μ is the mean and σ is the standard deviation).*

	F_0		SoE		EoE		η	
	μ	σ	μ	σ	μ	σ	μ	σ
Low(L1)	112	14	28	16	0.11	0.09	0.55	0.18
Medium(L2)	141	29	34	12	0.19	0.12	0.56	0.18
High(L3)	172	37	42	9	0.26	0.13	0.59	0.19

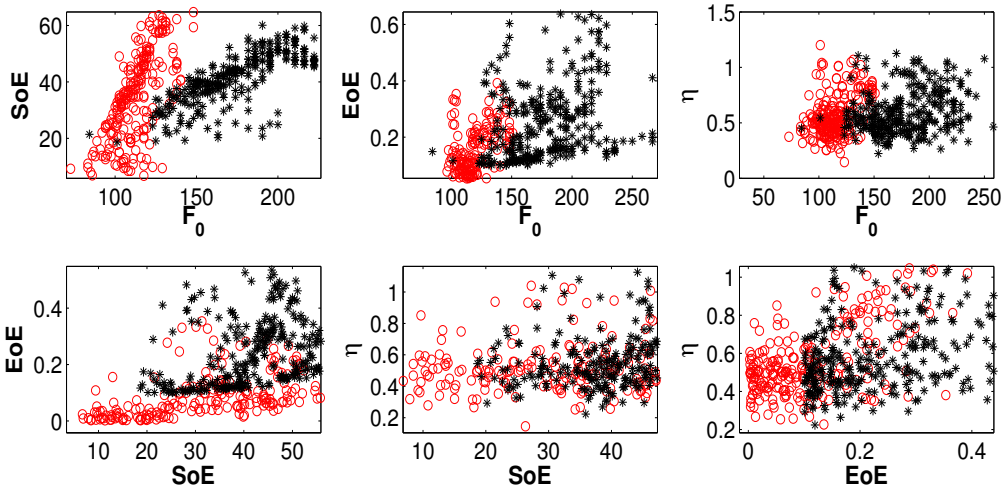

Figure 2: *The 2-D feature spaces for a speaker's reference (neutral) utterance (marked by 'o') and test (High Cognitive Load level) utterance (marked by '$*$').*

Table 4: *The average KL distance values among reference (Low load (L1)) utterance and test utterances of different cognitive loads (Low load (L1), Medium load (L2) and High Load (L3)) for Stroop test with time pressure task.*

	2-D feature spaces					
	C1	C2	C3	C4	C5	C6
L1 vs L1	0.270	1.064	0.045	1.921	0.081	2.105
L1 vs L2	0.528	2.014	0.412	1.816	0.310	4.260
L1 vs L3	0.413	4.120	0.095	2.426	0.184	3.190

Table 5: *The average KL distance values among reference (Low load (L1)) utterance and test utterances of different cognitive loads (Low load (L1), Medium load (L2) and High Load (L3)) for Stroop test with dual task.*

	2-D feature spaces					
	C1	C2	C3	C4	C5	C6
L1 vs L1	0.186	0.482	0.193	0.513	0.361	0.456
L1 vs L2	1.208	1.844	1.928	2.042	0.809	0.834
L1 vs L3	2.097	2.562	1.240	1.512	0.702	2.584

Table 6: *The average KL distance values among reference (Low load (L1)) utterance and test utterances of different cognitive loads (Low load (L1), Medium load (L2) and High Load (L3)) for reading span task.*

	2-D feature spaces					
	C1	C2	C3	C4	C5	C6
L1 vs L1	0.524	2.346	1.059	0.859	0.615	1.834
L1 vs L2	3.672	17.84	4.242	5.044	0.324	2.462
L1 vs L3	2.325	6.864	1.045	11.76	0.982	1.682

for cognitive load classification. Currently, work is underway on an attempt to use these excitation source features by normalizing the speakers and features for the development of cognitive load classification system.

5 Summary

In this study, the effect of excitation source features in the production of cognitive loaded speech are analyzed. Based on the analysis, it is observed that the excitation source features discriminate among the different cognitive load levels. It is also shown that the cognitive loaded speech deviates from neutral speech. The deviations are captured through 2-D feature spaces by computing the KL distance. From the present study, it can be concluded that, the excitation source features carry the production information of cognitive loaded speech, and hence they may be useful for cognitive load classification. This work can be extended by exploiting the other excitation source features and vocal tract system features.

References

Paavo Alku. 2011. Glottal inverse filtering analysis of human voice production-a review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana*, 36(5):623–650.

Hynek Boril, Seyed Omid Sadjadi, Tristan Kleinschmidt, and John H. L. Hansen. 2010. Analysis and detection of cognitive load and frustration in drivers' speech. In *INTERSPEECH*, pages 502–505.

Siyuan Chen, Julien Epps, and Fang Chen. 2011. A comparison of four methods for cognitive load measurement. In *OZCHI*, pages 76–79.

N. Dhananjaya and B. Yegnanarayana. 2010. Voiced/nonvoiced detection based on robustness of voiced epochs. *IEEE Signal Process. Letters*, 17(3):273–276, March.

T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit. 2012. Detection of glottal closure instants from speech signals: a quantitative review. *IEEE Trans. on Audio Speech and Language Processing*, 20(3):994–1006.

Thomas Drugman, Paavo Alku, Abeer Alwan, and Bayya Yegnanarayana. 2014. Glottal source processing: From analysis to applications. *Computer Speech & Language*.

Raul Fernandez and Rosalind W. Picard. 2003. Modeling drivers speech under stress. *Speech Communication*, 40(12):145 – 159.

P Gangamohan, Sudarsana Reddy Kadiri, and B. Yegnanarayana. 2013. Analysis of emotional speech at subsegmental level. In *INTERSPEECH*, pages 1916–1920, August.

M.Asif Khawaja, Fang Chen, Christine Owen, and Gregory Hickey. 2009. Cognitive load measurement from users linguistic speech features for adaptive interaction design. volume 5726 of *Lecture Notes in Computer Science*, pages 485–489.

Femke Kirschner, Fred Paas, and Paul A. Kirschner. 2009. Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25(2):306–314.

Phu Ngoc Le, Julien Epps, Eric H. C. Choi, and Eliathamby Ambikairajah. 2010. A study of voice source and vocal tract filter based features in cognitive load classification. In *ICPR*, pages 4516–4519.

Phu Ngoc Le, Eliathamby Ambikairajah, Julien Epps, Vidhyasaharan Sethu, and Eric H. C. Choi. 2011. Investigation of spectral centroid features for cognitive load classification. *Speech Communication*, 53(4):540–551.

Pisoni D. Van Summers W. Bernacki R. Lively, S. 1993. Effects of cognitive workload on speech production: acoustic analyses and perceptual consequences. *J. Acoust. Soc. Amer.*, 93:2962–2973.

J. Makhoul. 1975. Linear prediction: A tutorial review. 63:561–580, Apr.

K. Sri Rama Murty and B. Yegnanarayana. 2008. Epoch extraction from speech signals. *IEEE Trans. Audio, Speech, Lang. Process.*, 16(8):1602–1613, Nov.

K. Sri Rama Murty, B. Yegnanarayana, and M. Anand Joseph. 2009. Characterization of glottal activity from speech signals. *IEEE Signal Process. Letters*, 16(6):469–472, June.

Lysanne S. Post, Tamara van Gog, Fred Paas, and Rolf A. Zwaan. 2013. Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, 29(4):1450–1455.

Klaus R. Scherer, Didier Grandjean, Tom Johnstone, Gudrun Klasmeyer, and Thomas Bänziger. 2002. Acoustic correlates of task load and stress. In *INTERSPEECH*.

Björn Schuller, Stefan Steidl, Anton Batliner, Julien Epps, Florian Eyben, Fabien Ringeval, Erik Marchi, and Yue Zhang. 2014. The INTERSPEECH 2014 Computational Paralinguistics Challenge: Cognitive & Physical Load. In *INTERSPEECH*, Singapore, September. to appear.

Guruprasad Seshadri and B Yegnanarayana. 2009. Perceived loudness of speech based on the characteristics of glottal excitation source. *Journal of the Acoustical Society of America*, 126:2061–2071.

Hansen J.H.L. Steeneken, H.J.M. 1999. Speech under stress conditions: overview of the effect on speech production and on system performance. pages 2079–2082, Phoenix, AZ, USA. ICASSP.

Tet Fei Yap, Eliathamby Ambikairajah, Eric H. C. Choi, and Fang Chen. 2009. Phase based features for cognitive load measurement system. In *ICASSP*, pages 4825–4828.

Tet Fei Yap, Eliathamby Ambikairajah, Julien Epps, and Eric H. C. Choi. 2010a. Cognitive load classification using formant features. In *ISSPA*, pages 221–224.

Tet Fei Yap, Julien Epps, Eliathamby Ambikairajah, and Eric H. C. Choi. 2010b. An investigation of formant frequencies for cognitive load classification. In *INTERSPEECH*, pages 2022–2025.

Tet Fei Yap, Julien Epps, Eric H. C. Choi, and Eliathamby Ambikairajah. 2010c. Glottal features for speech-based cognitive load classification. In *ICASSP*, pages 5234–5237.

Tet Fei Yap, Julien Epps, Eliathamby Ambikairajah, and Eric H. C. Choi. 2011a. Formant frequencies under cognitive load: Effects and classification. *EURASIP J. Adv. Sig. Proc.*, 2011.

Tet Fei Yap, Julien Epps, Eliathamby Ambikairajah, and Eric H. C. Choi. 2011b. Voice source features for cognitive load classification. In *ICASSP*, pages 5700–5703.

T. F. Yap. 2012. *Speech production under cognitive load: Effects and classification*. Phd dissertation, The University of New South Wales, Sydney, Australia.

B. Yegnanarayana and Suryakanth V Gangashetty. 2011. Epoch-based analysis of speech signals. *Sadhana*, 36(5):651–697.

B. Yegnanarayana and K. Sri Rama Murty. 2009. Event-based instantaneous fundamental frequency estimation from speech signals. *IEEE Trans. on Audio, Speech, and Language Processing*, 17(4):614–624, May.

Bo Yin, Fang Chen, Natalie Ruiz, and Eliathamby Ambikairajah. 2008a. Speech-based cognitive load monitoring system. In *ICASSP*, pages 2041–2044.

Bo Yin, Natalie Ruiz, Fang Chen, and Eliathamby Ambikairajah. 2008b. Exploring classification techniques in speech based cognitive load monitoring. In *INTERSPEECH*, pages 2478–2481.

Bo Yin, Natalie Ruiz, Fang Chen, and Eliathamby Ambikairajah. 2008c. Investigating speech features and automatic measurement of cognitive load. In *MMSP*, pages 988–993.