2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

ICARUS: An Mwave† Based Real-time Speech Recognition System in Noise and Lombard Effect

Douglas A. Cairns and John H. L. Hansen
Digital Speech Processing Laboratory
Department of Electrical Engineering
Duke University
Durham, North Carolina 27706

10.21437/ICSLP.1992-238

1 Abstract

Automatic speech recognition by computer must address two issues to perform reliably in actual recognition environments. The first issue is real-time system performance. The second issue is the effect of background noise and/or speaker stress on recognition performance. A considerable effort has been made to develop speech recognition in tranquil environments. However, speech recognition algorithms formulated in tranquil environments generally perform poorly in adverse environments (background noise and/or speaker stress). In this paper, we propose a real-time recognition system called ICARUS which addresses the effect of background noise/speaker stress. The motivation for our stress compensation scheme is discussed with respect to the variation of speech characteristics spoken in noisy environments. Results of the proposed system are given for several speakers. ICARUS represents a first attempt at real-time speech recognition in adverse conditions. Early performance results, though not consistent over all speakers, show improvement in recognition of as much as +7.1 % for noise free Lombard speech, +15.7 % and +7.1 % for Lombard speech corrupted by additive white Gaussian and non-stationary cooling fan noise respectively.

2 Introduction

The advancement of high speed signal processing chip technology in recent years has significantly improved the prospects and needs for further speech recognition research. A wide assortment of recognition algorithms has been formulated for recognizing isolated words, phrases, and limited forms of continuous speech. Studies based on man/machine interaction and modality studies with computer input devices have shown voice to be the fastest, most natural form of communication for accomplishing complex tasks [1, 2]. In addition, by employing the modality of speech recognition, a users' hands and eyes may be used to accomplish other tasks.

Speech recognition in a tranquil laboratory environment does not address factors found in real recognition scenarios. Factors such as noise and speaker stress (due to the performance of a secondary task) can degrade speech and ultimately recognition performance. In addition to degrading the acoustic speech signal, noise causes the Lombard effect. The Lombard effect occurs because the speaker perceives background noise and changes their production system to communicate more efficiently in the noisy environment [3]. Speech recognition systems formulated in tranquil laboratory environments perform poorly in the presence of noise and speaker stress (Lombard effect). To address this problem, researchers have suggested recognition training and testing in actual noise environments. Results from this approach show that performance can be quite good if the training/recognition signal-to-noise ratio (SNR) is closely matched [4]. However, a mismatch of as little as 5dB can cause a significant degradation in recognition performance [5]. To address this, Hansen and Clements [6, 5, 7, 8] investigated tandem speech enhancement-stress compensation recognition algorithms. They demonstrated improvement in recognition performance on the average by +27 % (+42 % for Lombard speech) [9]. These results show that if issues of noise and stress are addressed, improved recognition performance can result.

Several issues impede the unification of these enhancement algorithms with existing recognition schemes for real-time applications. First, the enhancement algorithms require extensive computational resources. Second, they require knowledge of the type of stress (e.g., speech spoken under loud, angry, Lombard, etc. conditions). Finally, the stress compensation algorithms require knowledge of phoneme boundaries which is a challenging research problem given the presence of noise. For this study, only the Lombard stress condition be considered. In this paper, we formulate an algorithm for real-time speech recognition in noise and Lombard effect on a special purpose digital signal processing card. The algorithm is a speaker dependent, low vocabulary, isolated word speech recognition system. To illustrate the effects of noise on input speech, we first consider an analysis of speech signals and parameters in section 3. Next, section 4 presents algorithm formulation of the real-time recognition system called *ICARUS*. Finally, section 5 presents a recognition evaluation of ICARUS in additive white Gaussian noise and nonstationary cooling fan noise. Novel aspects of ICARUS include better speech parameterization techniques in noise, adaptive noise suppression prior to parameter extraction, and compensation of varying speech parameters caused by speaking in noise.

3 Analysis of Speech in Noise

Since the focus is speech recognition in noisy or adverse environments, information related to speech quality and/or speech parameters in noise will be useful in developing a robust recognition system. To address the quality of speech in noise, a study was performed to determine which sounds are most affected by increasing levels of noise. Speech (from the TIMIT/DARPA database [10]) was degraded with additive white Gaussian noise (AWGN) and non-stationary cooling fan noise from an IBM PS/2 computer at global signal-to-noise ratios (SNR) of 10, 20, and 30dB. The following objective measure was used to quantify speech quality[11].

$$d_{IS(j)}(\vec{a}_x, \vec{a}_{\hat{x}}) = 10 \cdot \log_{10}\left[1 + \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{A_{\hat{x}}(j\omega) - A_x(j\omega)}{A_x(j\omega)}\right| d\omega\right]. \quad (1)$$

Here, $d_{IS(j)}$ represents the Itakura-Saito distortion between the j^{th} original and noisy speech frames, $\vec{a}_x$, $\vec{a}_{\hat{x}}$ are the linear predictive coefficients from the j^{th} frame of the original and noisy speech waveforms respectively, and $\mathbf{R}_{\hat{x}}$ is the autocorrelation matrix of the noisy speech. Table 1 shows the results of speech quality measures for individual phonemes and measures grouped across sound classes. AWGN affects low energy plosives, most vowels, diphthongs, and liquids while PS/2 cooling fan noise primarily affects stops and fricatives. A speech recognition system that is perceptually based must address this loss in quality to perform reliably in noise.

To address the effects of speech spoken in noise, we consider a statistical analysis of two parameter classes to understand how speech parameters vary under the Lombard effect. These include linear predictive coefficients (LPC predictor and PARCOR coefficients), and mel-cepstral parameters. This determines noise sensitivity of parameters used to characterize speech, as well as those algorithms which estimate these parameters. Analysis was conducted on four data sets; i) noisefree neutral, ii) noisefree Lombard, iii) noisy neutral, and iv) noisy Lombard condition. Speech used for this analysis came from a data base previously established for analyzing speech under stress[12, 9]. Results show that approximately half of all the parameters resulted in

†Mwave is a trademark of the International Business Machines Corporation.

*This work sponsored in part by grants from the I.B.M. Corp. Grant No. N-UN-225-00, and the National Science Foundation Grant No. NSF-IRI-90-10536.

	ITAKURA-SAITO QUALITY MEASURES					
	White Gaussian			PS-2 Cooling Fan		
	10dB	20dB	30dB	10dB	20dB	30dB
Phoneme	**Individual Phoneme Measures**					
eh	1.989	0.306	0.020	0.012	0.001	0.000
er	4.368	0.815	0.075	0.041	0.002	0.000
oy	3.336	0.673	0.058	0.016	0.001	0.000
iy	3.665	1.557	0.285	0.450	0.025	0.001
axr	12.948	2.998	0.368	0.025	0.001	0.000
n	3.021	1.644	0.425	1.018	0.127	0.007
p	2.612	1.553	0.429	3.171	0.430	0.023
pcl	1.617	1.254	0.773	27.274	19.191	8.145
z	0.922	0.512	0.108	20.385	4.193	0.548
l	2.322	0.867	0.200	0.061	0.003	0.000
Sound Class	**Global Quality Measures over Sound Classes**					
Silence or Noise	3.311	2.733	1.629	15.122	11.343	5.526
Vowel	3.326	0.767	0.092	0.088	0.005	0.000
Nasal	3.021	1.644	0.425	1.018	0.127	0.007
Stops	3.056	2.060	0.924	12.769	8.216	3.365
Fricatives	1.202	0.540	0.089	14.878	2.860	0.348
Liquids and Glides	2.022	0.720	0.163	0.058	0.003	0.000
Voiced and Unvoiced	2.885	1.079	0.308	5.043	2.329	0.845
Overall	2.947	1.317	0.498	6.495	3.628	1.519

Table 1: Objective speech quality measures of phonetically balanced speech in white Gaussian and PS-2 cooling fan noise at three signal-to-noise ratios. Individual phoneme and global measures over sound classes are shown.

significant shifts in mean values in both noise free and noisy settings [13]. This suggests the need for some form of Lombard compensation for robust speech recognition in noise.

4 Algorithm Formulation

The analysis presented in the previous section shows that robust speech recognition in noisy environments must address the issues of noise and Lombard effect. It has been shown that noise cancellation and Lombard compensation significantly improve speech recognition in noise [14, 15]. In actual recognition environments such as a helicopter cockpit or an office, noise will be present as well as the real-time performance constraint. To address the problem of real time speech recognition in noise, the ICARUS speech recognition system was developed. ICARUS was implemented on an Mwave based digital signal processing (DSP) platform. The following sections discuss the DSP subsystem and give an overview of ICARUS.

4.1 Mwave Digital Signal Processing Subsystem

The Mwave subsystem proto-type card consists of a core DSP with external memory (64 Kbytes data; 96 Kbytes instruction) and peripheral hardware. The peripheral hardware performs I/O, DMA, and telephony functions. The core DSP is a Harvard architecture processor with a three phase pipeline. Each phase of the pipeline contains an instruction consisting of three operations that can be executed in parallel; read/write, ALU, and multiply. Given that this Mwave DSP is capable of up to 17 MIPS, the total possible throughput is 51 million operations per second. The Mwave data word is 16 bits, and fixed point arithmetic is used. A 16x16 multiplier is built into the DSP, thereby supporting 32 bit multiply and accumulate.

An operating system is provided which allows to efficiently schedule DSP resources. The operating system also provides multi-tasking environmental support. The multi-tasking environment is transparent to the developer as long as certain guidelines are followed with respect to data flow between tasks. The developer must ensure that sufficient processing power is available for the required tasks to operate in parallel.

4.2 ICARUS Algorithm Formulation

Starting with the real time system constraint, ICARUS was formulated so that all DSP processing steps require a limited window of speech (32-2048 samples). The flow diagram for ICARUS is shown in Figure 1. The system extracts mel-cepstrum parameters [16] from successive frames of speech (128 point analysis window, 50 % overlap), performs parameter data compression through vector quantization [17], and performs recognition using a hidden Markov model based algorithm [18, 14]. Speech is submitted to the system, where an endpoint detector determines the occurrence of speech activity. If no speech activity is present, the system characterizes background noise/silence. If speech activity is present, the system extracts mel-cepstral parameters and performs data compression via the vector quantizer. After performing data compression, the system determines if the endpoint detector has found the end of speech activity. If speech activity is complete, the system performs recognition. If speech activity is not complete, the system continues to gather coded speech parameters. The contributions of the ICARUS system are threefold: i) demonstrate automatic compensation of speech parameters caused by Lombard effect, ii) perform real-time speech recognition, and iii) automatically perform noise cancellation. The following sections will discuss the real time considerations for noise cancellation and Lombard compensation.

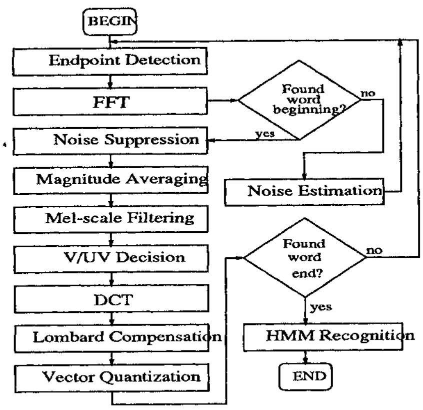

Figure 1: Flow diagram for speech recognition system

For a more complete discussion of the real time algorithm formulation, see [20, 13].

4.3 Noise Suppression

The ICARUS system incorporates a fairly standard approach to spectral subtraction with magnitude averaging to remove noise[19]. Spectral subtraction with magnitude averaging was chosen because the processing required only a small window of speech (real time constraint). To determine the level of noise removal, the average noise level is determined in the frequency domain and used for spectral subtraction.

The noise estimation procedure uses an average noise power spectrum as the estimate of the true noise power spectrum. When the endpoint detector indicates no speech activity, extracted frequency data is used for noise power spectrum characterization. An average noise power spectrum is estimated by averaging four consecutive noisy frames

$$|\hat{N}_{ave}(e^{j\omega})|^2 = \frac{1}{4}\sum_{i=1}^{4}|(N_i(e^{j\omega})|^2 \qquad (2)$$

where $N_i(e^{j\omega})$ is the frequency spectrum of the i^{th} noisy frame. The resulting noise power spectrum estimate is used in the spectral subtraction task portion of the system.

Spectral subtraction and magnitude averaging are frequency domain techniques. The assumption made is that noise reduction is possible by improving only the estimate of the magnitude spectrum. For ICARUS, power spectrum values are used in the standard spectral subtraction and magnitude averaging process in place of frequency magnitude values, due to the increased computational requirements of implementing a square root operation on the fixed point DSP.

4.4 Lombard Compensation

The ICARUS system uses an average weighting technique for Lombard compensation which was initially considered in [14]. This scheme was motivated by earlier stress compensation methods in noisefree and noisy settings [8]. As part of the HMM training procedure, average neutral mel-cepstral parameters for all tokens of each word are determined via

$$\overline{NM}(i,j) = \frac{1}{N_T}\sum_{j=1}^{N_T} MFCC(i,j) \quad 1 \le i \le 35, \quad 1 \le j \le 8 \qquad (3)$$

where

$\overline{NM}(i,j)$ - j^{th} average neutral mel-cepstral coefficient of the i^{th} word

$MFCC(i,j)$ - j^{th} mel-cepstral coefficient [16] of the i^{th} word

N_T - total number of frames over all tokens

i - word number

A similar procedure is used to calculate the average Lombard mel-cepstral coefficients $\overline{LM}(i,j)$. An average weighting vector for each word was then determined from

$$W(i,j) = \overline{NM}(i,j) - \overline{LM}(i,j). \qquad (4)$$

The weighting vector is then used to compensate the mel-cepstral coefficients according to

$$CM(i,j) = W(i,j) + MFCC(j) \quad \forall i \qquad (5)$$

where

$CM(i,j)$ - 35 compensated vectors of mel-cepstral coefficients

$W(i,j)$ - 35 weighting vectors

$MFCC(j)$ - vector of uncompensated Lombard mel-cepstral coefficients.

This procedure was computed on a frame by frame basis, which satisfies the real time constraint.

5 Evaluation

The main purpose for developing ICARUS is to address real-time speech recognition in adverse conditions. For our evaluations, four speaking conditions were considered; noise free neutral, noisy neutral, noise free Lombard, and noisy Lombard. The noise considered was 30dB of additive white Gaussian noise and 30dB of non-stationary cooling fan noise from an IBM PS/2 personal computer. An average noise level of 30dB may not appear significant, but it has been shown [13] that the actual noise level in consonant sections can be as low as 0-10dB for this noise level. Since consonants provide discriminatory power for confusable pairs, this is a reasonable noise level to consider.

The ICARUS system was evaluated using a previously established database [5, 9]). The database incorporated several different speakers talking under varying speech and stress conditions. For this study, two general male speakers were used in evaluating system performance. The speaking styles used were Lombard and neutral. Lombard effect was simulated by having speakers produce speech while listening to 85dB SPL of pink noise through headphones. The 35 word vocabulary for the neutral and Lombard speaking styles contained several confusable pairs ('six-fix', 'go-oh-no', 'white-wide') and many words with leading and trailing unvoiced fricatives or plosives ('point', 'strafe', 'freeze'). Such leading and trailing fricatives or plosives can cause errors for an endpoint detector in the presence of noise. From the 35 word vocabulary, a less confusable 10 word vocabulary was also chosen and used in system testing.

5.1 Recognition Performance

In all recognition evaluations, six of the 12 tokens of each of the 35 word vocabulary were used for training. All twelve tokens were used for recognition (i. e. all tests semi-open).

5.1.1 Noise-Free Performance

The noise-free performance of the ICARUS system was determined for both the neutral and Lombard speaking conditions. As shown in Table 3, the average neutral noise-free recognition rate was 91.7%. For the less confusable 10 word vocabulary (speaker 1), the recognition rate was 100.0% (see Table 4). These results are comparable to results obtained on a floating point workstation for a similar recognition system [12, 15]. For the 35 word vocabulary, results for Lombard speech show a decrease in recognition between 18-38 %. The addition of Lombard compensation improves the average recognition by 2.9% (7.1% for speaker 1), although improvement is not consistent over all speakers. Further studies employing more advanced non real-time workstation algorithms suggest that this performance could be improved if: i) VQ codebook size is increased to 128, especially under Lombard effect conditions, ii) training data for HMM and Lombard compensation vectors is increased, and iii) state-model sizes are increased to seven. The present formulation resulted in a 7.1% increase in recognition for one speaker, and no loss in performance for a second speaker. This Lombard compensation step therefore improves or at least maintain recognition performance. Recognition results for the 10 word vocabulary with Lombard speech shows a consistent recognition level respect to the 35-word vocabulary. Results for the 10 word vocabulary show a 10 % decrease in recognition. The addition of Lombard compensation decreases performance by an additional 30 %. We attibute this to the Lombard compensation reducing errors between confusable pairs rather than distinct words (see section 5.2.1. Since the 10 word vocabulary contained no confusable pairs, a performance decrease is not unreasonable. Also, it must be emphasized that there were limited tokens for training HMM models and compensation parameters.

5.2 Performance in AWGN and PS/2 Fan Noise

Performance of ICARUS in noise was determined by degrading speech two general male speakers with 30dB of AWGN and 30dB of PS/2 cooling fan noise. System performance was obtained with both noise suppression and Lombard compensation inactive, and then both activated. As indicated in Table 3, the 35 word vocabulary ICARUS system performance degrades significantly (between 52-55%) in the presence of Lombard condition and AWGN. An average of 9% improvement was obtained with Lombard compensation and noise cancellation, although improvement was seen to vary from speaker to speaker. Similar results were obtained with speech degraded with 30dB of PS/2 cooling fan noise. However, no consistent improvement was obtained with Lombard compensation and noise suppression. For the 10 word vocabulary (see Table 4), performance degrades by 20% for Lombard speech and AWGN. The addition of Lombard compensation did not improve recognition, but whole-word Lombard compensation and noise suppression did improve noisy Lombard performance by 5%. Reduction of vocabulary size and vocabulary difficulty increased ICARUS recognition performance by 18.6% in noisefree Lombard conditions, and 29.3% in noisy Lombard conditions.

5.2.1 Error Analysis

The vocabulary used to test the ICARUS system contained several confusable pairs, as well as words with similar, strong vowel sections ('go-oh-no','zero','hello'). It is suggested that these words would be substituted for one another under Lombard condition and cause errors. As can be seen from Figure 2, errors do result from confusable pairs and words with similar, strong vowel sections. The confusion matrices shown are for speaker G1 under noise-free Lombard condition. The first confusion matrix shows the results from testing the system without Lombard compensation. This matrix shows 'oh' being confused with 'hello' and 'zero' being confused with 'no'. The second confusion matrix shows the results from testing the system with Lombard compensation. This matrix shows a one-third reduction in the errors caused by confusable pairs and words with similar, strong vowel sections.

Results of System Testing for 35 Word Vocabulary					
Speaking Condition	Lombard Compensation	Noise Suppression	Recognition Rate (%)		
			Speaker 1	Speaker 2	Average
Neutral-noise free	off	off	92.4	91.0	91.7
Lombard-noise free	off	off	54.3	72.8	63.5
Lombard-noise free	on	off	61.4	71.4	66.4
Lombard-30dB AWGN	off	off	40.0	34.3	37.2
Lombard-30dB AWGN	on	on	41.4	50.0	45.7
Lombard-30dB PS/2	off	off	40.0	54.3	47.2
Lombard-30dB PS/2	on	on	47.1	42.8	44.9

Table 3: Performance of the ICARUS real-time speech recognizer in neutral, Lombard effect, and noisy Lombard effect speech conditions.

Results of System Testing for 10 Word Vocabulary			
Speaking Condition	Lombard Compensation	Noise Suppression	Recognition Rate (%) Speaker 1
Neutral-noise free	off	off	100.0
Lombard-noise free	off	off	90.0
Lombard-noise free	on	off	85.0
Lombard-30dB AWGN	off	off	70.0
Lombard-30dB AWGN	on	on	75.0

Table 4: Performance of the ICARUS real-time speech recognizer in neutral, Lombard effect, and noisy Lombard effect speech conditions.

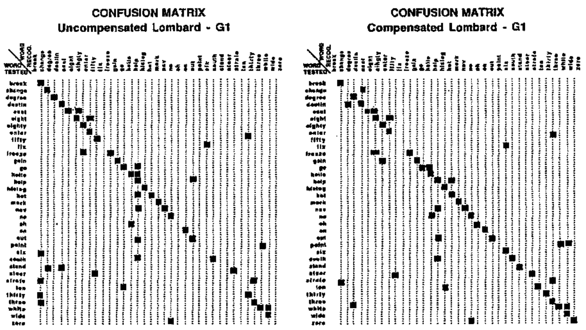

Figure 2: Confusion matrices for speaker 1 producing speech under Lombard effect, with and without Lombard compensation during recognition.

6 Conclusions and Future Work

Speech recognition outside of controlled laboratory conditions provides several challenges including; i) real-time system performance, and ii) the ability to perform reliably in the presence of noise. The long term goal of this research is to develop signal processing algorithms which efficiently address these challenges. The ICARUS system represents a platform for future work in this area. From our work on ICARUS, we can make several conclusions. First, real-time performance is comparable to non-real-time workstation implementation in noise free neutral conditions. Second, although noise suppression did improve recognition performance, further improvement is necessary if the system is to perform reliably in high noise environments. The specific noise suppression task is a much simplified version of one shown to improve recognition performance by 22 percent on a floating point workstation. Third, further improvement in the Lombard compensation phase is desirable, especially with regards to consistency of performance over speakers. ICARUS does demonstrate the real-time ability of integrated stress compensation, noise suppression tasks during speech recognition. Since a highly confusable vocabulary employing a limited number of training tokens was used, care should be exercised if comparisons are to be made with other developmental systems. In effect, this evaluation represents a "worst case" recognition scenario. Finally, the conservative reminder of the need for real time operation during algorithm formulation has left us with unused computation resources which can be used for growth and improvement. Currently about 33 percent of the Mwave DSP is utilized, leaving 67 percent of the computing resources for more robust noise cancellation and Lombard compensation.

For future work, in addition to improving the noise cancellation and/or Lombard compensation, we will investigate reformulating two key elements of ICARUS. The first area of investigation is the endpoint detector. The current endpoint detector has difficulty with leading and trailing unvoiced consonants. For a confusable vocabulary, this is a serious limitation. We have obtained encouraging preliminary results from a detection theory approach to boundary detection, and will incorporate this approach into ICARUS. The second area of investigation is the vector quantizer/ discrete HMM tandem. We believe that the vector quantizer has limited the effectiveness of the Lombard compensation. We are currently investigating removing the vector quantizer and replacing the discrete observation HMM with a semi-continuous hidden Markov model (SCHMM) approach. Our initial results based on the SCHMM approach have been promising.

References

[1] W.A. Lea, *Trends in Speech Recognition*, Pub. Prentice-Hall, Englewood Cliffs, NJ, 1980.

[2] *Readings in Speech Recognition*, Edited by A. Waibel, K.F. Lee, Morgan Kaufmann Publishers, Inc., San Mateo, Calif., 1990.

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

[4] B.A. Dautrich, L.R. Rabiner, T.B. Martin, "On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition", *IEEE Trans. on ASSP*, pg. 793-806, August 1983.

[5] J.H.L. Hansen, M.A. Clements, "Constrained Iterative Speech Enhancement with Application to Automatic Speech Recognition," *Proc. 1988 IEEE ICASSP*, pp. 561-564, April 1988.

[6] J.H.L. Hansen, M.A. Clements, "Iterative Speech Enhancement With Spectral Constraints", *Proc. 1987 IEEE ICASSP*, pp. 189-192, April 1987.

[7] J.H.L. Hansen, M.A. Clements, "Constrained Iterative Speech Enhancement with Application to Speech Recognition," *IEEE Trans. on Sig. Proc.*, pp. 795-805, April 1991.

[8] J.H.L. Hansen, M.A. Clements, "Stress Compensation and Noise Reduction Algorithms for Robust Speech Recognition," *Proc. 1989 IEEE ICASSP*, pp. 266-269, May 1989.

[9] J.H.L. Hansen, "Evaluation of Acoustic Correlates of Speech Under Stress for Robust Speech Recognition." *IEEE Proc. Northeast Bioeng. Conf.*, pp. 31-32, Boston, Mass., March 1989.

[10] "Getting Started With The DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database," National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, (prototype as of December 1988).

[11] S.R. Quackenbush, T.P. Barnwell, M.A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, 1988.

[12] J.H.L. Hansen, "Analysis and Compensation of Stressed and Noisy Speech With Application to Robust Automatic Recognition," Ph.D. Thesis, Georgia Inst. of Tech., July 1988.

[13] J.H.L. Hansen and D. Cairns, "ICARUS: Formulation of a Real-Time Speech Recognition System in Noise and Lombard Effect," in prep. for *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing.*

[14] J.H.L. Hansen, O.N. Bria, "Lombard Effect Compensation for Robust Automatic Speech Recognition in Noise," *Proc. of 1990 Inter. Conf. on Spoken Lang. Proc.*, pp. 1125-1128, Kobe, Japan, Nov. 1990.

[15] J.H.L. Hansen, O. Bria, "Improved Automatic Speech Recognition in Noise and Lombard Effect," *EURASIP-92, Sixth European Signal Proc. Conf.*, pp. S3.P-13.1-4, Heverlee, Belgium, Aug. 1992.

[16] S.B. Davis, P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. on ASSP*, pp. 357-366, Aug. 1980.

[17] R.M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, pp. 4-29, April 1984.

[18] S.E. Levinson, L.R. Rabiner, M.M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition," *Bell Sys. Tech. J.*, pp. 1035-1074, April 1983.

[19] S.F. Boll, " Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. on ASSP*, pp. 113-120, April 1979.

[20] D.A. Cairns, "Real-Time Speech Recognition Under Lombard Effect and in Noise," Master of Science Thesis, Dept. of Electrical Engineering, Duke University, May 1991.