UT-Scope: Speech under Lombard Effect and Cognitive Stress

Ayako Ikeno, Vaishnevi Varadarajan, Sanjay Patil, and John H.L. Hansen

The Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering & Computer Science
University of Texas at Dallas; Richardson, Texas 75083, USA,
{Ayako.Ikeno,John.Hansen}@utdallas.edu http://crss.utdallas.edu

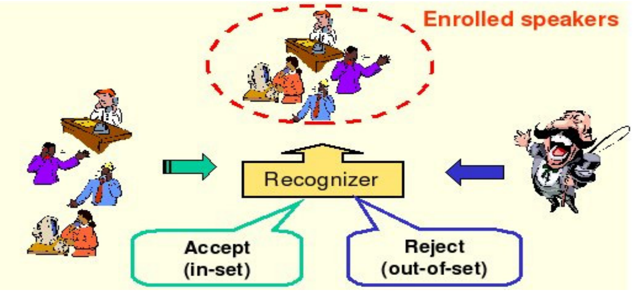
Abstract—This paper presents UT-Scope data base, and automatic and perceptual an evaluation of Lombard speech in In-Set Speaker Recognition. The speech used for the analysis forms a part of the UT-SCOPE database and consists of sentences from the well-known TIMIT corpus, spoken in the presence of highway, large crowd and pink noise. First, the deterioration of the EER of an in-set speaker identification system trained on neutral and tested with Lombard speech is illustrated. A clear demarcation between the effect of noise and Lombard effect on noise is also given by testing with noisy Lombard speech. The effect of test-token duration on system performance under the Lombard condition is addressed. We also report results from In-Set Speaker Recognition tasks performed by human subjects in comparison to the system performance. Overall observations suggest that deeper understanding of cognitive factor involved in perceptual speaker ID offers meaningful insights for further development of automated systems.[1][2][3]

TABLE OF CONTENTS

1. INTRODUCTION

Automatic Speaker recognition [1] plays an important role in the area of forensics and security as well as in speech communication such as recognizing a speaker for an automatic speech recognition or dialogue system.

Further development of *In-Set Speaker ID* system[2] is required for a variety of security related applications, such as monitoring individuals who belong to a defined group vs. who do not, as illustrated in Fig. 1.



In **Figure 1** speakers in the dotted circle represent In-Set speakers who should be accepted.

The advances in speech technology have led to an increased deployment of automatic speech systems in varying environments, such as factories, busy offices, cars, lecture halls, and wireless PDAs. This presents challenges to researchers dealing with a wide range of variability in speech characteristics, not only in acoustic and between-speaker variation but within speaker variation.

In this paper, we illustrate UT-Scope (Speech under Cognitive and Physical Stress and Emotion) Corpus and experimental results using speech under Lombard effect. The Lombard effect may be defined as speech produced due to increased vocal effort on the part of the speaker to improve the communication efficiency over environmental noise. It has been shown by Rajasekaran et al. [3] that Lombard effect degrades speech system performance to a greater degree than noise itself. Several approaches in the past have aimed at bridging the differences between training and testing conditions with respect to speech recognition systems[4,5,6,7]

Many studies have been initiated to analyze the characteristics of speech produced in noise[8,9,10,11]. These analyses have considered only individual words spoken under the Lombard effect whereas real speech systems use sentences as test utterances. In this paper, we perform our analyses on sentences. Further, previous studies have concentrated on speech produced under a single noise type. Since performance of speech systems vary according to noise type, one can expect the same for the Lombard effect also. Hence, compensation schemes for the Lombard effect should inherently depend on the nature of environmental noise.

The results from our perceptual experiments further show that Lombard speech contributes to In-Set Speaker ID performance but interferes with Out-of-Set speaker ID

[1] 1-4244-0525-4/07/$20.00 ©2007 IEEE.
[2] IEEEAC paper #1528, Version 2, Updated January 16, 2006
[3] This work was supported by RADC under contract No. FA8750-05-C-0029, and by University of Texas at Dallas under Project Emmitt.

detection. The trends also indicate that higher confidence ratings corresponds to higher accuracy when reference and test conditions match but not under mismatched conditions.

The following section describes UT-Scope Corpus.

2. UT-Scope

UT-Scope Corpus contains speech data under 4 types of stress: Lombard effect, cognitive stress, physical stress, and emotion. Data collection was performed in an ASHA certified sound booth, using a DAT recorder (FOSTEX) unit and three microphones – a Shure Beta, far-field desktop, and throat microphone.

2.1 Lombard effect

In previous studies, it has been shown that Lombard effect speech varies from neutral in terms of pitch, intensity, duration, spectral slope, formant location and bandwidth structure, etc.[10,13]. These differences cause a breakdown of speech system performance when systems are trained with neutral but tested with Lombard effect speech. In order to compensate for the variations in Lombard effect speech, it is meaningful to investigate the variations of Lombard effect under different noise types and noise levels. Previous research using SUSAS[13] have considered only a single noise type and level. Thus, a prior knowledge of the noise type and SNR would allow for more advanced and appropriate degree of compensation [14,15].

In this database, speech under three noise types at three levels is recorded. Pink noise (PNK), large crowd noise (LCR) and noise in a car traveling at 65 mph on a highway (HWY) with windows half open[16] are presented binaurally at different levels using open-air headphones worn by the speaker. This way, we provide a direct acoustic path for the subject speaking under Lombard effect, yet record a noise-free speech data sequence. The open-air headphones allow the speaker to hear his own voice when speaking as well. A pure-tone hearing test was performed for each speaker prior to data collection to objectively identify any potential hearing problems for the subjects under test. Speech samples from 59 speakers were collected. Noise presentation levels varied from 65 dB-SPL to 90 dB-SPL, representing 3 levels for each of the 3 noise types (i.e., 9 Lombard conditions).

The speech under Lombard effect consists of 100 phonetically-balanced read sentences chosen from the TIMIT corpus under neutral condition. 20 sentences, forming a subset of the aforementioned 100 sentences are used under each of the 9 Lombard effect conditions. The read speech also contains 5 tokens each of the 10 digits (0-9). These text materials were presented using a flat LCD display, with sentences presented in random order for every condition. Additionally, spontaneous speech of one minute duration is recorded by having the subject describe the content of visual images presented as part of the prompts.

2.2 Cognitive stress

Very often, one might access an automatic speech system while performing a cognitively demanding task like driving a car under heavy traffic conditions. Thus, studying the effect of cognitive stress on speech is of practical significance. In the UT-Scope corpus, speakers drive a car-driving simulator using a Sony PlayStation2 in scenarios that require extensive concentration. That is, the driving simulator has extreme cognitive task conditions. A standard size automobile steering wheel, and gas and brake pedals are used to perform the driving task. A driving seat which incorporates movements from the video player is used. The vibration effect which is transferred to the steering wheel adds to the reality of the simulator. In addition to the 3 microphones, video and biometrics such as heart rate and blood pressure are also recorded for this task. Speech samples from 60 speakers were collected.

2.3 Physical task stress

Physical stress includes factors such as G-force experienced in aircraft cockpits, stress experienced due to high speeds in racing cars etc. In this corpus, speech is collected while a person operates a stair stepper. Video and biometrics data are also recorded for this task as well. Speech collection from 60 speakers was completed.

2.4 Emotion

Speech with emotions such as anxiety, fear and anger is common when accessing automatic speech systems. For example, a person trying to access his bank account on his cell-phone might get frustrated due to repeated failures of the voice-based security system. Speech under emotion will also be represented in UT-Scope corpus (we have previously employed the Soldier of the Month paradigm [17] in our algorithm development for stress detection and assessment). This part of data will be collected in the spring 2007.

3. In-Set Speaker ID performance

3.1. System description

The speech data collected under different Lombard effect conditions were tested on an in-set speaker ID system. An in-set speaker ID system is one that identifies if the speech input belongs to one of the group of speakers defined in the system. This back-end system employs a Universal Background Model (UBM) constructed from a selected set of speakers. A speaker specific MAP adapted Gaussian Mixture Model (GMM) is obtained from the UBM for each of the trained in-set speakers. The scores obtained by comparing the test utterances with the trained speaker models were normalized and thresholds were set using unconstrained cohort normalization likelihood ratio testing [18]. Further details of the GMM-UBM system can be found in [4]. Equal error rates (EER) were obtained using the in-set speaker ID system for the different Lombard effect conditions.

3.2. Experimental setup

3.2.1. Speaker and development set

A set of 30 speakers was chosen for the test set. The population consisted of 19 females and 11 males. 15 were in-set and the other 15 were out-of-set. Out of the 15 in-set speakers, 9 were females and 6 males. There were 10 female and 5 male out-of-set speakers. The development set consisted of 60 speakers chosen from the TIMIT corpus. The male-female ratio in the development set was maintained the same as the in-set speakers.

3.2.2. Front-end processing

Speech from all speakers was windowed with a Hamming window of 20ms duration with 10ms overlap rate. A 23-dimensional feature vector consisting of 19-dimensional MFCC's and 4 spectral center of gravity coefficients was extracted from all the speech data [19].

3.3. Experiments and results

Two sets of experiments were performed on the speaker ID system. Both experiments used training data consisting of ~30s (10 sentences) of neutral speech. The first set of tests investigated the degradation caused by Lombard effect only. The neutral-trained speaker ID system was tested with clean neutral and noise-free Lombard speech. The effect of test utterance duration was also investigated by using two sets of test utterance length, 3s and 12s. The results are shown in Tables 1 and 2.

Noise Type	*Noise Level 1*	*Noise Level 2*	*Noise Level 3*
HWY	**23.16**	**32.67**	**34.83**
LCR	**25.83**	**29.5**	**30.33**
PNK	**22.17**	**25**	**31.5**

Table 1 shows **EER (%)** of In-set Speaker ID System using **3 sec. clean** test utterances. EER with Neutral speech is 14.67%. Noise types are highway (HWY), large crowd (LCR), and pink (PNK) noise.

Noise Type	*Noise Level 1*	*Noise Level 2*	*Noise Level 3*
HWY	**20**	**29.5**	**34**
LCR	**24.5**	**30.17**	**28.83**
PNK	**16.8**	**22.16**	**31.5**

Table 2 shows **EER (%)** of In-set Speaker ID System using **12 sec. clean** test utterances. EER with Neutral speech is 7.2%

These results clearly show that Lombard speech degrades the performance of a speaker ID system. The average increase in the EER for under the different Lombard conditions with 3s test tokens is 93%, relative to the EER under neutral condition and that in the 12s test case is 266%. The absolute values of the EER show that an increase in the test duration helps in improving the EER under the neutral condition by about 50%, but the average improvement under the Lombard conditions is only 7.68%. Also, with increased test duration, EER reduction under the highest level (Level 3) of noise is negligible (2.4 %). Hence, increased test duration does not improve the EER under Lombard conditions and therefore, the Lombard effect changes the spectral structure to the point where additional test material cannot recover the performance.

The second set of experiments was performed by degrading the neutral and Lombard speech test tokens. However, the training was done with clean neutral speech only. This was considered in order to determine if speaker ID performance is more significantly impacted by noise type/level, or speech production changes due to Lombard effect. The noise used for producing the Lombard conditions was used for degrading the respective utterances. The SNRs used for large crowd, highway and pink noise conditions were 5dB, -5dB and 0 dB respectively. These noise levels represent the exact noise present when collecting noise-free Lombard speech. The experiments were repeated for 3s and 12s test utterances. The results are summarized in Tables 3 and 4.

Noise Type	*Noisy NEU*	*Noise Level 1*	*Noise Level 2*	*Noise Level 3*
HWY	**49.33**	**53.33**	**54.167**	**54.167**
LCR	**46.33**	**48.33**	**53**	**51.5**
PNK	**48.33**	**48.99**	**48**	**50.167**

Table 3 shows **EER (%)** of In-set Speaker ID System using **3 sec. degraded** test utterances. Clean Neutral EER is 14.67%

Noise Type	*Noisy NEU*	*Noise Level 1*	*Noise Level 2*	*Noise Level 3*
HWY	**45.49**	**51.33**	**52.67**	**56.17**
LCR	**42.5**	**49.5**	**49.3**	**50**
PNK	**49.33**	**44.67**	**52.16**	**51.5**

Table 4 shows **EER (%)** of In-set Speaker ID System using **12 sec. degraded** test utterances. Clean Neutral EER is 7.2%

The first column in the above tables marked NOISY NEU represent the EER for neutral speech degraded with the respective noise types (i.e. noisy speech without the Lombard effect). It is evident from that the Lombard effect along with noise degrades system perform more than noise only. Also, we can see that the error rates are not additive, in the sense that the EER with noisy neutral and clean Lombard speech do not sum up to the EER with noisy Lombard speech. Here, it is noticeable that the increase in the test duration does not help at high Lombard levels (Level 3). Lombard speech with noise clearly results in very poor performance. When speech enhancement algorithms for noisy Lombard speech do not address Lombard effect, we can only move up to the performance shown for clean Lombard speech in Tables 4 and 5. Achieving true effective performance for speech enhancement in noisy Lombard speech therefore requires normalization of the Lombard effect [20].

4. LISTENER TEST: EXPERIMENTAL SET UP

The listeners for this experiment were drawn from students at the University of Texas at Dallas. Thirty listeners were native English speakers and 17 were nonnative. Nonnative speakers' native languages were Chinese(2), Hindi(8), Korean(2), Spanish(1), Thai(1), Turkish(1), Urdu(1), and Vietnamese(1). All the listeners reported no history of hearing loss or problems.

To conduct a set of perceptual experiments, the speech samples were extracted from UTScope[8] corpus, described in Section 2. Lombard speech used in this perceptual experiment was produced while the speaker listened to highway driving noise at 90dB-SPL through open-air head phones. Each speech sample in this experiment was composed of 3 read phonetically-balanced sentences. Read speech was selected for this study since the speech is comparable among different speakers.

The listener test was conducted in an ASHA certified single-wall sound booth, using an interactional computer user interface and Bose^{TM} noise canceling headphones. The listeners were instructed to listen to each training/reference file before taking the test. The reference files for the 12 in-set speakers were accessible any time during the test as well.

Listeners were instructed to listen to each test file up to three times to determine whether the test speech was produced by one of 12 in-set speakers (IN) or someone else (OUT), and to indicate the confidence of their selection (1=not sure at all, 2, 3=somewhat sure, 4, or 5=absolutely sure). A total of 12 test speech samples were presented to each listener under each of the following three conditions:

NL-LD: *Mismatched condition.* Training/reference files contained neutral speech. Lombard speech was used as test audio.

LD-LD: *Matched condition.* Both training/reference and test files contained Lombard speech.

NL-NL: *Matched condition.* Both training/reference and test files contained neutral speech.

Eight of the test samples were In-Set speakers and 4 were Out-of-Set speakers. The statistical analyses are performed based on the three speech conditions (NL-LD, LD-LD, and NL-NL), using repeated measures ANOVA.

In reporting the results in this section, we employ the following terms: (a) *Accuracy* – in-set speakers are correctly identified as in-set, or when out-of-set speakers are correctly identified as out-of-set, (b) *False Reject* – in-set speakers are incorrectly identified as out-of-set speakers, and (c) *False Accept* – out-of-set speakers are incorrectly identified as in-set speakers. Those rates were calculated simply as a ratio. For example if 7 out of 8 in-set speakers in the test set were recognized correctly by a listener, the accuracy rate is 87.5%.

5. Perceptual Speaker ID Results

In this section, we report experimental results from the listener test. The average accuracy shows that the effect of the conditions (NL-LD, LD-LD, NL-NL) on perceptual In-Set speaker ID is significant (p<.0001). With the mismatched condition, listener performance is significantly lower (NL-LD: Native:57%, Nonnative:53%) compared to the matched conditions (LD-LD: 78%, 67%, NL-NL: 71%, 70%), as shown in Fig. 2. This trend agrees with the system performance described in Section 3.

Unlike language related tasks, such as accent or dialect ID[24], listeners' language background does not show significant effect on perceptual In-Set speaker ID. Overall, native vs. nonnative listeners' performance does not show a significant difference. However, in the case of the LD-LD condition, native listeners' accuracy is noticeably higher (78%) than nonnative listeners' accuracy (67%).

The following subsections describe the analysis of In-Set speaker ID and Out-of-Set speaker ID results as well as confidence ratings.

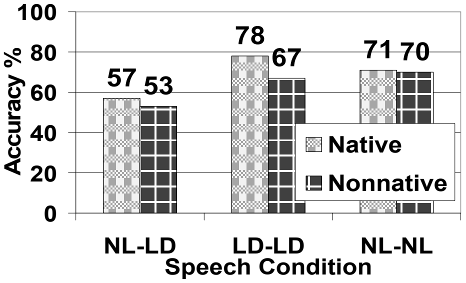


Figure 2 shows **average accuracy** of perceptual in-set speaker ID using neutral (NL) and Lombard (LD) speech.

5.1 In-Set Results

For In-Set speaker ID results, the effect of the speech condition is significant (p<.0001). As shown in Figures 3 and 4, In-Set speaker detection accuracy is significantly higher with the matched conditions (LD-LD and NL-NL) compared to the mismatched condition (NL-LD). With the mismatched condition, false reject rate is high (Native:50%, Nonnative:58%). It is also important to note that the accuracy is significantly higher with Lombard speech (LD-LD: 88%, 79%) than with neutral speech (NL-NL: 73%, 75%) for both listener groups (p=.0036).

Listeners' language background (native vs. nonnative) does not show statistical significance. However, it should also be noted here that in the cases of NL-LD and LD-LD, the native listeners' accuracy for In-Set speaker ID is noticeably higher compared to the accuracy of nonnative listeners.

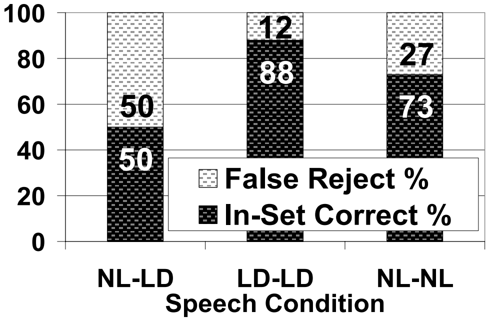


Figure 3 illustrates **native listeners**' performance with **In-Set speakers** using natural (NL) and Lombard (LD) speech.

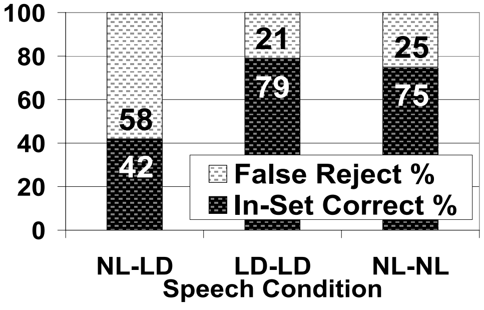


Figure 4 illustrates **nonnative listeners**' performance with **In-Set speakers** using natural (NL) and Lombard (LD) speech.

5.2 Out-of-Set Results

In the case of Out-of-Set speakers as well, the effect of speech condition on perceptual ID is shown to be significant (p=.0001). With the Out-of-Set speakers, accuracy is significantly higher under the mismatched condition (NL-LD: Native:71%, Nonnative:75%), compared to the matched conditions, as illustrated in Figures 5 and 6. On the other hand, when the reference and test conditions match, false accept rate is high for both native and nonnative listener groups, especially with Lombard speech (41%, 57%), compared to NL-NL.

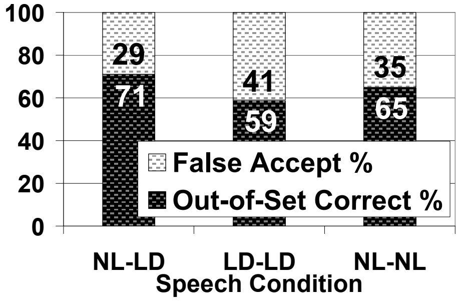


Figure 5 shows **native listeners**' performance with **Out-of-Set speakers** using natural (NL) and Lombard (LD) speech.

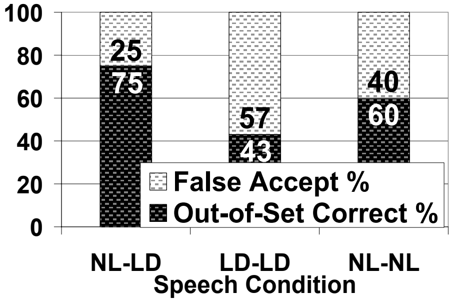


Figure 6 shows **nonnative listeners**' performance with **Out-of-Set speakers** using natural (NL) and Lombard (LD) speech.

Taken together with the In-Set results (i.e., high false reject rate for NL-LD), those trends indicate that, when the reference and test speech conditions do not match, listeners tend to perceive speakers as out-of-set. In addition, considering the high ID accuracy for LD-LD and NL-NL in the case of In-Set results, our observations also suggest that, when the reference and test speech conditions match, listeners tend to perceive speakers as in-set.

5.3 Confidence Ratings and Accuracy

In addition to the ID performance results, confidence ratings also show significant effect of speech conditions (p<.0001), as shown in Table 5. The ratings for the mismatched condition are significantly lower (3.5, 3.7) than for the matched conditions (LD-LD: 4.1, 4.1, NL-NL: 4.0, 3.9). This suggests that confidence measures are somewhat relevant to the accuracy scores.

Further analysis on the accuracy and token coverage[4] based on confidence ratings indicate the following trends, as illustrated in Fig. 7: (i) when reference and test conditions match (LD-LD, NL-NL), the higher the confidence rating, the higher the accuracy (LD-LD shown on the right in Fig. 7), (ii) when reference and test conditions do not match (NL-LD), confidence ratings and accuracy do not show consistent relation (on the left in Fig. 7), and (iii) token coverage decreases significantly with mismatched conditions (18% at confidence 5) compared to matched conditions (LD-LD:42%, NL-NL:35% at confidence 5).

	NL-LD	LD-LD	NL-NL
Native	3.5	4.1	4.0
Nonnative	3.7	4.1	3.9

Table 5 shows **confidence ratings** on average for native and nonnative listeners. The confidence was rated between 1 (not sure at all) and 5 (absolutely sure), as described in Section 2.

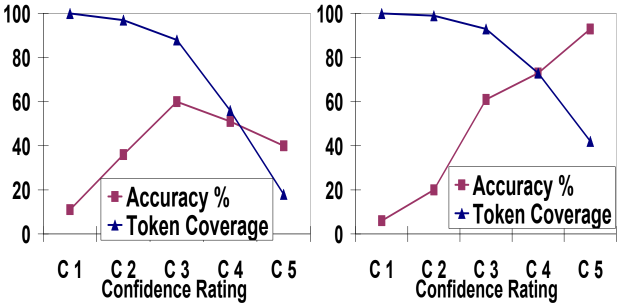


Figure 7 shows **native listeners**' **accuracy** and **token coverage** based on **confidence ratings** (C1–C5). Left is ***NL-LD.*** Right is ***LD-LD***.

6 AUTOMATIC VS. PERCEPTUAL SPEAKER ID

This section illustrates comparison of automatic system and human performance. In Varadarajan and Hansen[21], we further report results from In-Set speaker ID(15in/15out). The trends show some similarity between automated system performance and human performance. In the case of the automatic system, as illustrated in Table 5, EER is significantly higher with the mismatched condition (NL-LD: 36.33%). It is also the case that EER is lower with Lombard speech (LD-LD: 9.66%) compared to neutral speech (NL-NL: 11.67%).

	NL-LD	LD-LD	NL-NL
EER %	36.33	9.66	11.67

Table 5 shows **Equal Error Rate** results for **In-set/out-of-set speaker ID** performed by an **automated system**.

The numbers are not directly comparable between human performance and machine performance due to various differences, such as the amount of training and test data, or numbers of speakers used in the experiments. However, it is meaningful to notice similar trends between human and machine behavior, since this suggests that deeper understanding of human perception can contribute to

[4] Token coverage is the amount of listener responses to be included.

the further advancement of automated systems. For example, transformations which attempt to imitate the effects of stressed speech can be used to synthesize stressed speech from neutral speech [22] in order to improve the performance of automated systems.

More detailed analysis on the listener test results indicated that, when training/reference and test conditions match, Lombard speech impacts listener perception in a way that contributes to higher In-Set speaker detection accuracy but reduces Out-of-Set speaker detection accuracy, compared to neutral speech condition.

7. CONCLUSIONS

This study has considered an analysis of the characteristics of speech under different types of Lombard conditions. I was shown that EER of in-set speaker ID system shows degradation under Lombard effect. It was also shown that increasing test duration does not improve the system performance under the Lombard conditions, indicating fundamental changes in phoneme spectral structure. In addition, compensation for only noise under noisy Lombard conditions keeps the system performance far from baseline performance. This represents the first study to investigate the change in speech production for Lombard effect under different noise types and levels. Our overall observation indicates that while noise impacts speaker ID performance, speech production under Lombard effect causes fundamental changes in spectral structure for a GMM that cannot be overcome by simply using longer test sequences.

The results from perceptual speaker recognition also showed that speaker ID accuracy is significantly lower when reference and test data do not match. The trends also showed that Lombard speech contributes to higher accuracy in In-Set speaker ID, but decreases correct detection of Out-of-Set speakers. Furthermore, the analysis indicated that confidence ratings correspond to accuracy when reference and test conditions match but not under mismatched conditions. Taken together with the performance from automated systems, overall observations point to the importance of further investigation on cognitive aspects involved in speaker recognition, which will contribute to development of combined automatic-human based systems as well as stand-alone automatic systems.

REFERENCES

[1] D.A. Reynolds, "An overview of automatic speaker recognition technology", ICASSP, 2002.

[2] Angkititrakul, P. and Hansen, J.H.L. "Discriminative In-set/Out-of-set Speaker Recognition," IEEE Trans. Speech & Audio Processing, (to appear in Feb 2007).

[3] Rajasekaran, P., Doddington G., and Picone, J. "Recognition of speech under stress and in noise," Proc. ICASSP, 1986.

[4] Angkititrakul et al., "Cluster-dependent Modeling and Confidence Measure Processing", ICSLP 2004.

[5] Chen Y, "Cepstral domain talker stress compensation for robust speech recognition", IEEE Trans. ASSP 36, April 1988.

[6] Stanton B.J. et al. "Robust recognition of loud and Lombard speech in the fighter cockpit environment", ICASSP 89.

[7] Hansen J.H.L. "MCE-ACC for Speech Recognition in Noise and Lombard Effect", IEEE Trans. SAP, vol.2, Oct.1994.

[8] Summers W. et al. "Effects of noise on speech production: Acoustical and perceptual analyses", JASA, vol. 84, 1988.

[9] H.J.M. Steeneken, J.H.L. Hansen, "Speech Under Stress Conditions: Overview of the Effect of Speech Production on Speech System Performance," ICASSP 99.

[10] Hansen J.H.L., "Analysis and compensation of stressed and noisy speech with application to robust automatic recognition", Thesis, Georgia Inst. Tech. July 1988.

[11] Stanton B.J. et al. "Acoustic-Phonetic analysis of loud and Lombard speech in simulated cockpit conditions", ICASSP 88.

[12] Junqua J. "The Lombard reflex and its role on human listeners and automatic speech recognizers", J. Acoust. Soc. Amer., Jan. 1993.

[13] J.H.L. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communications,* 20(2), 151-170, 1996.

[14] J.H.L. Hansen, S. Bou-Ghazale, "Getting Started with SUSAS: A Speech Under Simulated and Actual Stress Database," *EUROSPEECH-97*, vol.4, p.1743-1746, Rhodes, Greece, Sept. 1997.

[15] J.H.L.Hansen,"Morphological Constrained Enhancement with Adaptive Cepstral Compensation (MCE-ACC) for Speech Recognition in Noise and Lombard Effect," *IEEE Transactions on Speech & Audio Processing*, 2(4), pp. 598-614, Oct. 1994.

[16] M. Akbacak, J.H.L Hansen, *Environmental sniffing: noise knowledge estimation for robust speech systems*, International Conference on Acoustic and Speech Signal Processing 2003.

[17] E. Ruzanski, J.H.L Hansen, et.al, *Improved "TEO" feature-based automatic stress detection using physiological and acoustic speech sensors* , Interspeech 2005.

[18] Fortuna J., Sivakumaran P. et al. "Open-set speaker identification using adapted Gaussian mixture models", Interspeech 2005.

[19] Hansen J.H.L. et al. "Constrained Iterative Speech Enhancement with Application to Speech Recognition", IEEE. Trans. Sig. Proc. 1991.

[20] Varadarajan, V., Hansen, J.H.L., and Ikeno, A. "UTScope - A corpus for Speech under Cognitive/Physical task Stress and Emotion. Language Resources and Evaluation (LREC) 2006.

[21] Ikeno, A., and Hansen, J.H.L. "Perceptual Recognition Cues in Native English Accent Variation: Listener Accent, Perceived Accent, and Comprehension," Proc. ICASSP, 2006.

[22] Varadarajan, V., and Hansen, J.H.L. "Analysis and normalization of Lombard speech under different types and levels of noise with application to in-set speaker ID system," (submitted to IEEE Transactions).

BIOGRAPHY

John H.L. Hansen, (IEEE S'81-M'82-SM'93) received the Ph.D. and M.S. degrees in Electrical Engineering from Georgia Institute of Technology, Atlanta, Georgia, in 1988 and 1983, and B.S.E.E. degree from Rutgers University, College of Engineering, New Brunswick, N.J. in 1982. He joined University of Texas at Dallas, Erik Jonsson School of Engineering and Computer Science in the fall of 2005, where he is Professor and Department Chairman of Electrical Engineering, and holds the Distinguished University Chair in Telecommunications Engineering. He also holds a joint appointment as Professor in the School of Brain and Behavioral Sciences (Speech & Hearing). At UTD, he established the Center for Robust Speech Systems (CRSS) which is part of the Human Language Technology Research Institute. Previously, he served as

Department Chairman and Professor in the Dept. of Speech, Language and Hearing Sciences (SLHS), and Professor in the Dept. of Electrical & Computer Engineering, at Univ. of Colorado Boulder (1998-2005), where he co-founded the Center for Spoken Language Research. In 1988, he established the Robust Speech Processing Laboratory (RSPL) and continues to direct research activities in CRSS at UTD. He is serving as IEEE Signal Processing Society Distinguished Lecturer for 2005/06, Member of the IEEE Signal Processing Society Speech Technical Committee and Educational Technical Committee, and has served as Technical Advisor to U.S. Delegate for NATO (IST/TG-01), Associate Editor for IEEE Trans. Speech & Audio Processing (1992-99), Associate Editor for IEEE Signal Processing Letters (1998-2000), Editorial Board Member for the IEEE Signal Processing Magazine (2001-03). He has also served as guest editor of the Oct. 1994 special issue on Robust Speech Recognition for IEEE Trans. Speech & Audio Proc. He has served on the Speech Communications Technical Committee for the Acoustical Society of America (2000-03), and is serving as a member of the ISCA (Inter. Speech Communications Association) Board (2004-07). His research interests span the areas of digital speech processing, analysis and modeling of speech and speaker traits, speech enhancement, feature estimation in noise, robust speech recognition with emphasis on spoken document retrieval, and in-vehicle interactive systems for hands-free human-computer interaction. He has supervised 33 (17 PhD, 16 MS) thesis candidates, was recipient of the 2005 University of Colorado Teacher Recognition Award as voted by the student body, and author/co-author of 222 journal and conference papers in the field of speech processing and communications, coauthor of the textbook *Discrete-Time Processing of Speech Signals*, (IEEE Press, 2000), co-editor of *DSP for In-Vehicle and Mobile Systems* (Springer, Vol. 1 - 2004, Vol. 2 - 2006), and lead author of the report “The Impact of Speech Under ‘Stress’ on Military Speech Technology,” (NATO RTO-TR-10, 2000). He also organized and served as General Chair for ICSLP-2002: International Conference on Spoken Language Processing, Sept. 16-20, 2002, and will serve as Technical Program Chair for IEEE ICASSP-2010, Dallas, TX.