Getting Started with the SUSAS: Speech Under Simulated and Actual Stress Database

John H.L. Hansen
Sahar E. Bou-Ghazale
Ruhi Sarikaya
Bryan Pellom

Robust Speech Processing Laboratory
Duke University, Department of Electrical Engineering
http://www.ee.duke.edu/Research/Speech
E-Mail: jhlh@ee.duke.edu

April 15, 1998

Technical Report: RSPL-98-10
Release 1.4

ABSTRACT

It is well known that the introduction of acoustic background distortion and the variability resulting from environmentally induced stress causes speech recognition algorithms to fail. In this paper, we discuss SUSAS: a speech database previously collected for analysis and algorithm formulation of speech recognition in noise and stress. The *SUSAS* database refers to *Speech Under Simulated and Actual Stress*, and is intended to be employed in the study of how speech production and recognition varies when speaking during stressed conditions. SUSAS consists of five domains, encompassing a wide variety of stresses and emotions. A total of 32 speakers were employed to generate in excess of 16,000 isolated-word utterances. The five stress domains include: i) talking styles[1] (slow, fast, soft, loud, angry, clear, question), ii) single tracking task or speech produced in noise (Lombard effect), iii) dual tracking computer response task, iv) actual subject motion-fear tasks (G-force, Lombard effect, noise, fear), v) psychiatric analysis data (speech under depression, fear, anxiety). A common highly confusable vocabulary set of 35 aircraft communication words make up the data base (e.g., /go-oh-no/, /wide-white/, etc). All speech tokens were sampled using a 16-bit A/D converter at a sample rate of 8kHz. Simulated speech under stress data consists of data from ten stressed styles (talking styles, single tracking task and Lombard effect domains); while actual speech under stress data consists of speech produced while performing either (i) dual-tracking workload computer tasks, or (ii) subject motion-fear tasks (subjects in roller-coaster rides). Additional speech was also later added from four pilots in Apache helicopter flight conditions. This paper will discuss (i) the formulation of the SUSAS database, (ii) organization of stressed speaking styles, and (iii) summarize some of the previous studies which have used SUSAS data. SUSAS is presently available on CD-ROM through the NATO RSG.10.

[1] Approximately half of the SUSAS data base consists of style data donated by MIT Lincoln Laboratory (Lippmann, et al. [37], Chen [8], Hansen [19]).

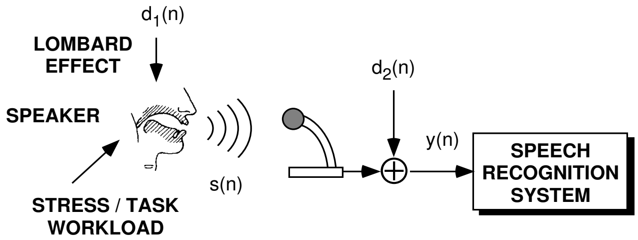

Figure 1: Types of distortion which can be addressed for robust speech recognition.

1. INTRODUCTION: Why Speech Recognizers Break

The issue of robustness in speech recognition can take on a broad range of problems. A speech recognizer may be robust in one environment and inappropriate for another. In Fig. 1, a general speech recognition scenario is presented which considers a variety of speech signal distortions. For this scenario, we assume that a speaker is exposed to some adverse environment, where ambient noise is present and a stress induced task is required. The adverse environment could be a noisy automobile where cellular communication is used, noisy helicopter or aircraft cockpits, noisy factory environments, and others. Since the user task could be demanding, the speaker is required to divert a measured level of cognitive processing, leaving formulation of speech for recognition as a secondary task.

Workload task stress has been shown to significantly impact recognition performance [8, 19, 23, 42, 47]. Since background noise is present, the speaker will experience the *Lombard* effect [40, 33]; a condition where speech production is altered in an effort to communicate more effectively across a noisy environment. In addition, a speaker may also experience situational stress, (i.e., anger, fear, other emotional effects), which will alter the manner in which speech is produced. If we assume $s(n)$ to represent a *Normal*, noise-free speech signal, then the acoustic signal at the microphone will include distortion due to stress, *Lombard* effect, and additive noise. Other distortions such as microphone mismatch, cellular/telephone channels, or voice coding effects can also degrade the speech signal. Therefore, the *Neutral* speech signal $s(n)$, having been produced and transmitted under adverse conditions, is transformed into the degraded signal $y(n)$.

$$y(n) = \left\{ \left[s(n) \middle| \begin{array}{l} \text{WORKLOAD TASK} \\ \text{STRESS} \\ \text{LOMBARD EFFECT } \{d_1\} \end{array} \right] + d_1(n) \right\} + d_2(n)$$

Having outlined the basic structure of a degraded speech signal under stress and noise, we now turn to the formulation of the SUSAS database. This database was designed for analysis and modeling of speech under stress in order to improve the robustness of speech recognition algorithms. Section 2 presents an overview of the structure of the SUSAS database. A discussion of the stress styles and workload tasks are also included. Section 3 summarizes some details on the database notation, speakers, and other organizational issues. Finally, Section 4 presents a brief summary of previous research studies which have used data from the SUSAS Speech Under Stress database.

2. Structure and Format of SUSAS

SUSAS represents a comprehensive speech under stress database which was formulated and collected by Hansen, 1988[19]. The database is partitioned into five domains, encompassing a wide variety of stresses and emotions. A total of 32 speakers (13 female, 19 male), with ages ranging from 22 to 76 were employed to generate in excess of 16,000 utterances. Four additional male speakers were later added in 1993 which included pilots flying missions in Apache helicopters. Fig. 2 illustrates the various domains present in the database. Each domain will be discussed separately in the following subsections. While it is of interest to obtain speech under actual stressful conditions, it is not always possible to obtain this in noisefree situations. However, some forms of cognitive or mild affective stress[2] are possible with the help of interactive workload computer tasks. As such, two of the domains employ computer response tasks. The four domains available in the present release of SUSAS consists of a 35 aircraft communication word vocabulary which is summarized in Table 1.

2.1 Talking Styles Domain

The first portion of the SUSAS Database involves speech under various speaking styles. In an earlier investigation, Lippmann, Mack and Paul [36, 43] considered a multi-style training procedure for a traditional hidden Markov model (HMM), speaker dependent, isolated word recognition system. The data used in these studies was originally sampled at 16kHz [36, 43], and was donated by Lincoln Laboratory (R. Lippmann and C. Weinstein). This portion of SUSAS contains utterances under eight speaking styles (normal, slow, fast, soft, loud, question, clear enunciation, angry). The vocabulary consists of 35 aircraft communication words containing a number of subsets that are difficult for recognition systems. These subsets include {go, hello, oh, no}, {six, fix}, {white, wide, point}, {degree, three, thirty, freeze}, and {eight, eighty, gain, change}. The 35 words make up a subset of 105 words which have been used by Texas Instruments to evaluate recognition systems. The words were produced by nine male talkers sampling three major dialects (General American, Boston, New York). Each word was produced 28 times by each subject for a total of 8,820 words.

2.2 Single Tracking Task Domain

The second domain of the SUSAS database consists of speech data from a single tracking task, and speech under Lombard effect. Here, the same same 35-word vocabulary and 9 speakers from talking styles domain produced speech while performing a computer workload task which stimulates stress originally proposed by Jex [31]. Jex formulated a set of standardized subcritical tasks for tracking workload calibration. The important aspect of this work was that graded levels of mental workload were possible. In this approach, a single tracking task is developed which is the response of a marginally stable, single-pole system. The operator views a display of the error between the command input and plant output, and corrects these with opposite pressure on a control stick (similar to a joy-stick in some video games). The degree of instability may be adjusted for varying degrees of difficulty. Two levels of workload difficulty were used in this section. Subjective ratings, performance data, and heart rate data indicated that the high workload (lambda = 70%) was measurably more difficult than the moderate level (lambda = 50%). This portion consists of a total of 1,890 isolated words.

This domain also includes a small portion of speech produced in noise, simulating the Lombard effect [40]. The Lombard effect occurs when talkers vary their speech characteristics in order to increase intelligibility when speaking in a noisy environment. For this portion, Pink noise was presented binaurally at an overall level of 85 dB SPL.

2.3 Dual Tracking Task Domain

The third SUSAS domain consists of speech produced while performing a dual (compensatory and acquisition) tracking task as a means of inducing workload. Here, the method proposed by Jex[31] was not considered, since it represents a single compensatory tracking task. Such a task does not approximate the demanding pilot stress in an aircraft cockpit. Therefore a dual tracking task which addresses the pilots' two key goals, flight control and target acquisition, was considered. Task difficulty could be controlled by time constraints for completion, or increasing resource competition or motivation. In this portion, a dual tracking task was

[2] We note here that *affective stress* is produced whenever the input to consciousness, whether it be from an immediate awareness of external events or a recollection of personally significant events of the past, is seen as threatening to the individual's safety, self esteem, or satisfaction of desires.

SUSAS Database

Speech under Simulated and Actual Stress

Domain	Type of Stress or Emotion	Speakers	Count	Vocabulary
Talking Styles	Simulated Stress: slow, soft, fast, loud, angry, clear, question	9 Speakers (all male)	8820	35 aircraft communication words
Single Tracking Task	Calibrated Workload Tracking Task: moderate & high stress; Lombard effect	9 Speakers (all male)	1890	35 aircraft communication words
Dual Tracking Task	Acquisition & Compensatory Tracking Task: moderate & high stress	8 Speakers (4 male) (4 female)	2257	35 aircraft communication words
Actual Speech under Stress	Amusement Park Roller-Coaster; Helicopter Cockpit Recordings (G-force, Lombard effect, Noise, Fear, Anxiety)	11 Speakers (4 male, 3 female) (4 male)	1642	35 aircraft communication words
Psychiatric Analysis	Patient Interviews: (Depression, Fear, Anxiety, Angry)	8 Speakers (6 female) (2 male)	600	Conversational Speech: Phrases & Sentences

Figure 2: SUSAS: Speech Under Simulated and Actual Stress database.

35-Word SUSAS VOCABULARY SET					
brake	eighty	go	nav	six	thirty
change	enter	hello	no	south	three
degree	fifty	help	oh	stand	white
destination	fix	histogram	on	steer	wide
east	freeze	hot	out	strafe	zero
eight	gain	mark	point	ten	

Table 1: A summary of the 35-word vocabulary set used for SUSAS domains 1 to 4 (talking styles, single tracking task, dual tracking task, actual speech under stress).

considered which was similar to that developed by Folds, Gerth and Engelman [12, 13] for the USAF School of Aerospace Medicine. The primary tracking task was a pursuit task in which the input signal was determined by the sum of two sine functions. A constant was added and subtracted from the function to form two parallel sinusoids which were displayed, giving the appearance of a winding road which was scrolled down a computer screen over time. The response marker (output signal) was a small circle. The vertical position of the circle was fixed at the center of the display; the horizontal position was determined by the movement of a control stick in its x-axis. Fig. 3 illustrates what an operator saw on his computer screen during the dual tracking task. The trail of response markers (triangles and circles) indicates past attempts by the operator to perform the task. The upper portion of the display represents the next few seconds of the task. As time progresses, the upper portion will scroll down towards the center of the display, requiring the operator's response. Therefore, the operator can to a certain extent, anticipate the primary tracking task as shown on the right half of the screen.

As the roadway moves downward, the operator's must adjust the control stick to position the circle as close to the center of the road as possible. After 20 seconds, a target acquisition task appeared on the left portion of the screen. Here, two narrowly spaced vertical lines were drawn along with a small triangle. The vertical position of the triangle was also fixed at the center of the display. A Gaussian distributed random value was added to the triangle's horizontal position which moved it to the left or right. The operator used the x-axis movement of a second control stick to position the triangle back to center of the two lines. The noise values were added at fixed times to the triangle's position, so the operator was never certain if movement at any time was a result of his actions or random movements. This represents the noise associated with an automatic target acquisition system which must be corrected by the pilot (e.g., automatic camera system). After 40 seconds of performing both tasks, the primary pursuit task was disabled, leaving the secondary target task active for the final 20 seconds.

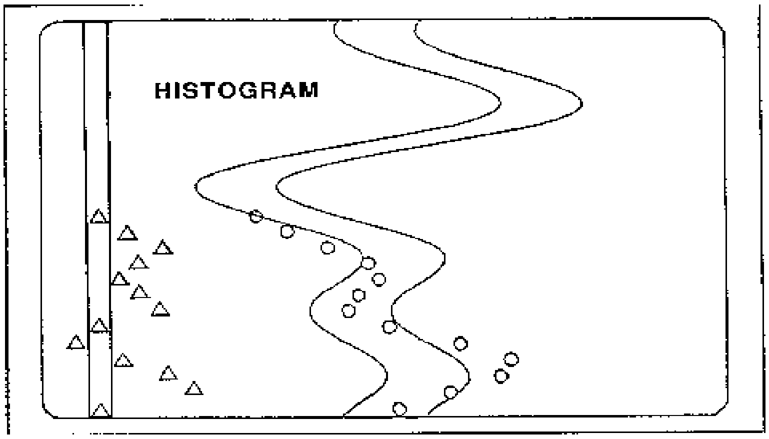

Figure 3: An example of what an operator would see during the performance of the dual tracking task. The primary tracking task on the right represents flight control, the secondary task represents target acquisition.

Several parameters were available for increased task difficulty, the simplest being the overall time allowed to perform the task. Other possibilities exist, however variation in overall time constraint was chosen in order to keep the basic task structure similar to the previously used and accepted workload procedure.

To facilitate an organized manner for collection of an operator's speech, randomized words from the 35 word list used in first two domains of SUSAS were displayed on the screen during all three stages. Each word appeared in the same position, and the operator was instructed to read them quickly while performing the dual tracking task. Operators wore a high quality head-mounted directional microphone for recording (Shure-512 unidirectional close-talking dynamic microphone). All recordings were performed using a Sony WM-D6C Professional Walkman using high quality metal oxide tape. When a new word appeared on the screen, a low frequency tone was emitted. In some cases, a variation in the length of time a word remained on the screen was also used in order to generate higher stress levels. For the moderate stress level, every three stage test lasted for 80 seconds. Each operator performed the three stage dual workload task nine times, each with a different type of Gaussian variation for the target acquisition task. A weighted RMS error was found for each task individually, and combined. This was displayed between tests so operators could observe how well they performed the tasks. Operators were instructed to give equal emphasis to all three tasks (pursuit task, target acquisition, and word entry). For the high workload case, the operator was required to perform all three tasks in half the time (40 seconds). Combined (primary and secondary) average weighted RMS errors for moderate and high dual tracking tasks for the eight speakers are shown in Fig. 4. The average over all speakers is also shown. A comparison of RMS errors indicate that the high workload case was significantly more taxing then the moderate case. Average error differences between moderate and high stress over nine trials for each speaker ranged from 0.5% to 9.5%. Speaker #1 and #6 resulted in similar average RMS errors for both moderate and high stressed conditions.

2.4 Actual Speech Under Stress

The fourth domain of SUSAS consists of speech produced during the completion of two types of subject motion-fear tasks. In order to attempt a simulation of the sudden change in altitude or direction which might be experienced in an aircraft cockpit, two types of motion tasks were considered. These tasks were chosen because they required no training, yet generated the type of stress (fear or anxiety) which might be experienced in an emergency situation. Two rides from an amusement park[3] were chosen as suitable, the *Scream Machine* and *Free Fall*. The *Free Fall* ride lasts for about 60 seconds, with the free fall portion comprising about 10 seconds. Four seated passengers are strapped in an upright seated position into a car which is raised vertically to approximately 130 feet. The car is moved forward where it pauses for several seconds before being released. It drops vertically downward for about 100 feet, before rolling onto a horizontal portion of the track for deceleration. During the free fall portion, talkers repeated several pre-chosen words from the 35 word list. Speech was recorded using a high quality head-mounted microphone and cassette recorder unit strapped to the talker's body.

The second motion task considered was the *Scream Machine*. This is a typical wooden frame roller-coaster which seats roughly 30-36 passengers. Due to the large number of passengers, higher levels of background screaming can be heard in these recordings. The overall ride consists of large vertical movements with small amounts of lateral movement during calm periods between drops. Initial tests gave little indication of variation in tape speed due to motion of the recording equipment. The entire ride lasts for about 90 seconds. Talkers were instructed to say the word *top* when their car reached the top of a hill. Speakers repeated

[3]Six-Flags Over Georgia, an amusement park located in the metropolitan Atlanta area.

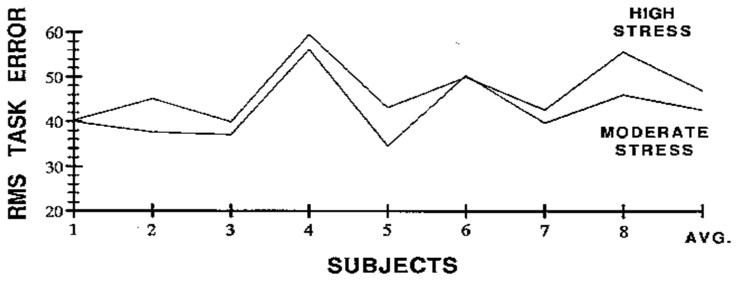

Figure 4: The combined (both primary and secondary tasks) average weighted RMS errors for moderate and high dual tracking tasks over eight speakers are shown (Female speakers: #1,#3,#5,#7, Male speakers: #2,#4,#6,#8). The average (over 72 trials per speaker) for both moderate and high task loads over all speakers is also shown.

words from a 35-word vocabulary card held in their hands. Each speaker (3 female, 4 male) performed the task twice. Due to the increased task time for this ride, larger amounts of stressed utterances could be obtained. The speaker's location during the ride was identified based on timing and background noise. Fig. 5 gives an overview of the ride, and illustrates how each recording was partitioned and subjectively marked for stress with respect to time and position during the task. The chosen subjects were all native Americans with no apparent speech deficiencies. A total of 1642 utterances were collected from speech under task stressed, and baseline neutral recordings for the seven speakers. In each subject motion task, at least four factors contributed to the type of speech recorded: g-force variation, background noise, Lombard effect, fear and/or anxiety.

While SUSAS was organized and collected during the period 1985-88, (Hansen, 1988[19]), it was determined that additional speech under actual stress would be useful for analysis and modeling in speech recognition. Four additional male speakers were later added in 1993 which included pilots flying missions in Apache helicopters. Two of the pilots were operating their Apache helicopter normal flight conditions as (i) baseline with helicopter on the ground but running, and (ii) pilots flying their helicopter while speaking. Both pilots used the same 35-word vocabulary set used in previous SUSAS domains. An additional set of recordings were included from two other Apache helicopter pilots flying a night mission into the Raleigh/Durham Airport while running low on fuel (the pilots were not familiar with the area, since they were from another state in the U.S.). This speech consists of tactical communications between pilot, co-pilot and sometimes an air-control person giving directions. The speech data consists of continuous speech passages not from the 35-word vocabulary. Stress levels increase as the fuel level begins to drop. Two large digitized files are included which contains speech from each pilot and co-pilot.

2.5 Psychiatric Analysis Domain

The last domain of the database is the psychiatric analysis domain. In general, the identification of emotion in speech has been an important facet for psychiatric analysis. A collection of recordings were obtained from Emory Medical University's Department of Psychiatry for the purposes of obtaining examples of speech under emotional stress. Patients undergoing psychiatric analysis were recorded using a high quality microphone and tape recorder in a natural doctor-patient environment. Recordings from eight patients (six female, two male) were obtained. Based on informal listening sessions, the overriding emotion present was mild to severe depression. In some cases, brief passages of fear, anxiety, and/or anger were also identified. The analysis of each recording began by carefully screening each tape and subjectively marking phrases or individual words as being under stress. Each utterance was then lowpass filtered at 3.7 kHz and sampled at 8 kHz. From these recordings, excellent examples were found of speech under mild to severe depression . At the present time, a vocabulary of well over 600 words or phrases has been collected. We note that at the time of this release, we decided not to include this data at this time until we could determine if permission was needed to release patient speech.

3. DATABASE NOTATION & ORGANIZATION

In this section, we summarize details on the database notation, speakers, and other organizational issues. The main directory of SUSAS contains the following three sub-directories:

`documentation/  simulated/  actual/`

The `documentation/` directory contains a postscript version of this documentation, and two readme files, one for each speech under stress sub-directory. We will summarize the Simulated (Section 3.1) and Actual (Section 3.2) speech under stress directories now.

3.1 Simulated Speech Under Stress Data

The `SUSAS/simulated/` directory contains speech from 9 speakers, which make up the Talking Styles and Single Tracking Task domains of SUSAS. The 9 speakers consist of three groups of speakers with (i) general USA accent (g1, g2, g3), (ii) New England/Boston, MA accent (b1, b2, b3), and (iii) New York City, NY accent (n1, n2, n3). Speakers were asked to produce

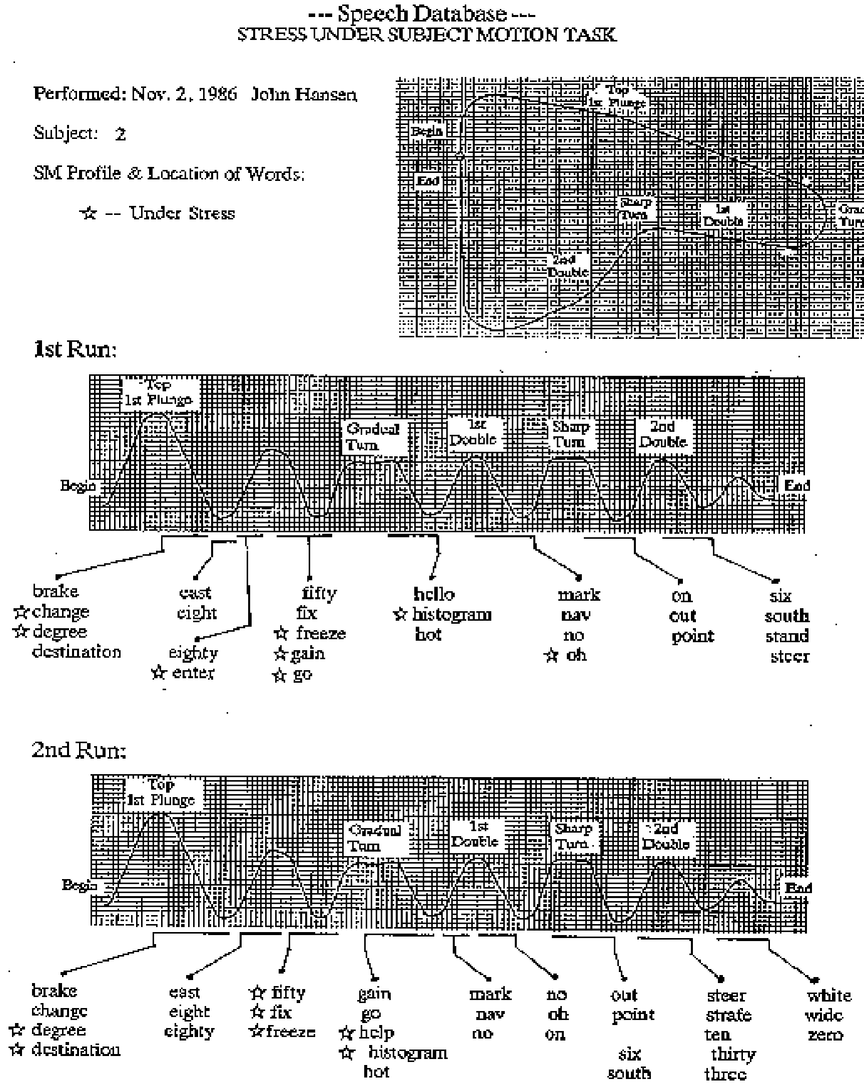

Figure 5: Profile of the *Scream Machine* motion task. The position of utterances were marked based on timing, and those under stress were extracted for analysis.

speech in a particular style while seated in a quiet environment. Fig. 2 summarizes the simulated speech under stress data conditions contained in the directory (sample extensions use speaker g2).

For each speaker and stress condition, a sub-directory of 70 words exists which contains 2 tokens each of the 35-word vocabulary. For example, in the directory `SUSAS/simulated/g2/angry`, you will find the files `break1.g2a ... zero2.g2a,` which represents the first and second tokens of each of the 35 words spoken by speaker G2 under the angry stress condition (i.e., extension `.g2a`). All speech files are sampled at an 8kHz sample rate, and represented using 16-bit integers (i.e., binary short format). Excessive leading and trailing silence has been removed via endpoint detection.

Finally, given the token count for each stress condition and speaker under simulated stress conditions, there should be 1,190 tokens per speaker under SUSAS/simulated. Unfortunately, a very small number of errors and missing files were present in the data originally collected by Lincoln Laboratory. In order to make sure that 2 tokens exist for each stress condition, one of the following three steps were taken:

(1) if one token exists for a given stress condition, this file was duplicated to make up the second token: (e.g., break1.b1n did exist, but break2.b1n was missing: therefore break1.b1n was copied to break2.b1n).

(2) if neither token existed for a given stress condition, then tokens from the neutral directory were used to make up both tokens: (e.g., for speaker g2, under slow conditions, both destin1.g2s and destin2.g2s did not exist: therefore destin1.g2n and destin2.g2n were copied to destin1.g2s and destin2.g2s).

(3) if a file was originally labeled incorrectly, we used the correct file when possible; otherwise the second stressed speech token is used in it's place: (e.g., the file on1.b1c7 was actually an original token of the

SUMMARY SIMULATED STRESS CONDITIONS		
Stress Condition	Number of Tokens per Stress Condition [35-word vocabulary] × Number.Tokens.Per.Condition	Word token Extension
angry	2 tokens of simulated anger	.g2a
clear	2 tokens of clearly enunciated speech	.g2c
cond50	2 tokens of computer workload, low task stress	.g2c5
cond70	2 tokens of computer workload, high task stress	.g2c7
fast	2 tokens of fast speech	.g2f
lombard	2 tokens of speech produced with 85dBA pink noise	.g2lom
loud	2 tokens of loudly spoken speech	.g2l
neutral	2 tokens of neutral speech	.g2n
question	2 tokens of speech asked as a question	.g2q
slow	2 tokens of slowly uttered speech	.g2s
soft	2 tokens of soft or whispered speech	.g2w
train	12 tokens of neutral training speech	.g2t

Table 2: Summary of the Directory structure from the `SUSAS/simulated/` directory.

word "oh" (most likely, the speaker said the wrong word): therefore on1.b1c7 was renamed on1.b1c7-OH, and on2.b1c7 was copied to on1.b1c7).

For the simulated stress portion of SUSAS, a total of 51 errors were addressed. A summary of the modifications are included in the file: `ReadMe-Simulated.txt` which can be found in the `SUSAS/documentation/` directory.

3.2 Actual Speech Under Stress Data

The `SUSAS/actual/` directory contains speech from 11 speakers, which make up the Dual Tracking Task and Actual Stressed Speech (speech from roller-coaster rides and helicopter cockpits) domains of SUSAS. All speech files from these domains of SUSAS were sampled at 8kHz, with samples represented as 16-bit integers.

There are 7 speakers who participated in both Dual Tracking tasks and Subject-Motion Fear (roller-coaster) tasks. The 4 male speakers have the following file extensions: m1, m2, m3, m4; and the 3 female speakers have these file extensions: f1, f2, f3. A sub-directory exists for each of these 7 speakers. Under each speaker directory are five stress domain directories (sample extensions use speaker m1) as shown in Table 3.

The vocabulary was taken from the 35-word set listed in Table 1. For neutral speech, recordings were made in a quite, sound resistant chamber (3-5 tokens of each of word was recorded in random order). A limited number of speech tokens were obtained from the Free Fall Amusement Park ride stress condition, since this task lasts for only 10 seconds. Also contained in each speaker directory are files which contain the complete speech from each ride, as well as files which contain samples of the background silence/noise obtained from within the recordings. For example, for the speaker m2, the file `m2/ff/All1.m2f` contains one complete run of the free-fall ride task. The file `m2/sm/All1.m2s` contains speech from one complete run of the scream machine roller coaster ride task. Files which are entitled `m2/sm/Silence1.m2s` represent sample silence/noise files which were extracted from the original recording. Finally, while speakers were instructed to produce speech from the 35-word set, they did at times produce other speech samples (this was especially true for the scream machine roller coaster ride). These words or phrases have also been digitized and stored in each speakers directory. For example, speaker m2 during the roller coaster ride task produced the words "pilot" and "may-day". These can be found in `m2/sm` as `pilot1.m2s` and `mayday1.m2s`. For these out-of-vocabulary words, neutral tokens were obtained and stored in the neutral speaker directory (e.g., in `m2/neutral` you will find `pilot1.m2n` and `mayday1.m2n`).

In the speaker directories `actual/rs` and `actual/dt` is speech from two Apache helicopter pilots (one pilot, one co-pilot) from the North Carolina National Guard. For these speakers, there are two sub-directories which contain: `actual/rs/medst/` baseline recordings of pilot and co-pilot producing the 35-word vocabulary with engines on, and the helicopter on the run-way; and `actual/rs/hist/` which contains the same pilot and co-pilot producing speech while flying their helicopter in a basic maneuvers (hover, turns, etc.). Also contained in each speaker directory are files named `actual/rs/All1.rsh` which represent full audio recordings of the warm-up and flight tasks for pilot and co-pilot.

Finally, the directory `actual/fuelout` contains speech from an Apache helicopter pilot and co-pilot during a night mission where they are running low on fuel. The pilot and co-pilot were flying from outside the state of North Carolina towards the National Guard base in North Carolina. During this speech, you will hear the pilot, co-pilot, and control tower attempting to give information on landmarks so the pilot could find the base. As the fuel level begins to run low, the pilot and co-pilot exhibit higher levels of stress. They finally locate the base and land safely. Speaker directory `actual/fuelout` contains two files `dwrf1` and `dwrf2` which contains speech from speakers `dw` and `rf`. This speech contains continuous speech of pilot, co-pilot, and control tower with discussions on finding the run-way (e.g., example speech includes discussion on landmarks such as 'powerlines', 'treeline', etc.).

3.3 Label Data for Speech Under Stress Data

In SUSAS Rev. 1.4, label files have been included for all speech data. The directory `/SUSAS_labs` contains a

SUMMARY ACTUAL STRESS CONDITIONS		
Stress Condition	Token Number varies per Stress Condition [35-word vocabulary]	Word token Extension
neutral	Neutral Speech (5 tokens)	.m1n
medst	medium level Dual-Tracking task stress (3 tokens)	.m1m
hist	high level Dual-Tracking task stress (3 tokens)	.m1h
ff	Free Fall: Amusement Park ride stress (3 tokens)	.m1f
sm	Scream Machine: Roller Coaster stress (3 tokens)	.m1s

Table 3: Summary of the Directory structure from the `SUSAS/actual/` directory.

directory structure which is similar to the speech data directory `/SUSAS`. The parsing routine was based on a new hidden Markov model based approach[44, 45],[4]. An example of the label file is given below for the parsed speech file *help1.g1a* spoken by speaker *g1* under *angry* speaking conditions (this is the file "help1.g1a.lab").

```
0     940   sil
940   1620  hh
1620  1900  eh
1900  2460  l
2460  2660  p
2660  2940  sil
```

Here, we see that silence samples occur for the first 940 samples, followed by the phonemes for the word help in order. Table 4 summarizes the phoneme set used for phonetically marking the SUSAS speech corpus.

4. SUMMARY of KEY SUSAS REFERENCES

In this section, we present a brief summary of previous research studies which have used data from the SUSAS Speech Under Stress database. To illustrate the problem of speech recognition in stress and noise, a baseline speech recognizer (VQ-HMM) was employed on noisefree and noisy stressed speech from SUSAS. This system was a discrete observation, 5-state, left-to-right HMM, trained in a speaker dependent mode (trained using odd neutral tokens from the `/train/` directory; and tested using even tokens (then repeated with even training, and odd token testing). A 64-state speaker-dependent VQ codebook was trained using two minutes of training data. Table 5 shows that when stress and noise are introduced, recognition rates decrease significantly. When white Gaussian noise is introduced, noisy stressed speech rates varied, with an average rate of $Avg10 = 30.3\%$ (i.e., a 58% decrease from the 88.3% neutral rate). Recognition performance also varies considerably across stressed speaking conditions as reflected in the large standard deviation in rate of recognition. ($StDev10 = 15.35, 9.12$ for noise free and noisy stressed conditions).

4.1 Stressed Speech Analysis

Below is a summary of the papers which have considered analysis of SUSAS stressed speech data. Note that some of these studies were focused on only simulated data, and others used actual stressed speech results to confirm simulated findings.

[1-anal] J.H.L. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communications*, Special Issue on Speech Under Stress, vol. 20, pp. 151-173, Nov. 1996.

[2-anal] J.H.L. Hansen and B.D. Womack, "Feature Analysis and Neural Network based Classification of Speech under Stress," *IEEE Trans. Speech & Audio Proc.,* vol. 4, no. 4, pp. 307-313, July 1996.

[3-anal] J.H.L. Hansen, "A Source Generator Framework for Analysis of Acoustic Correlates of Speech Under Stress. Part 1: Pitch, Duration, and Intensity Effects," submitted to *Journal Acoust. Soc. of America,* 44 pgs, Dec. 1995.

[4-anal] D. Cairns, J.H.L. Hansen, "Nonlinear Analysis and Detection of Speech Under Stressed Conditions," *Journal Acoust. Soc. of America,* vol. 96, no. 6, pp. 3392-3400, Dec. 1994.

[5-anal] D. Cairns, J.H.L. Hansen, "Nonlinear Speech Analysis using the Teager Energy Operator with Application to Speech Classification under Stress," *ICSLP-94: Inter. Conf. on Spoken Lang. Proc.,* vol. II, vol. 3, pp. 1035-1038, Yokohama, Japan, Sept. 1994.

[6-anal] J.H.L. Hansen, "Evaluation of Acoustic Correlates of Speech under Stress for Robust Speech Recognition." *IEEE Proc. Fifteenth Annual Northeast Bioengineering Conf.,* pp. 31-32, Boston, Mass., March 1989.

[7-anal] 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., Atlanta, GA, 428 pgs., July 1988.

[8-anal] J.H.L. Hansen, M.A. Clements, "Evaluation of Speech under Stress and Emotional Conditions," *Proc. Acoust. Society of America,* H15, vol. 82, Fall Supplement pp. S17, Nov. 1987.

4.2 Stressed Speech Synthesis

The following papers have considered methods for imparting stress onto neutral input speech.

[1-syn] S. Bou-Ghazale, J.H.L. Hansen, "Stressed Speech Synthesis Based on a Source Generator Framework," Speech Communications: Special Issue on Speech Under Stress, vol. 20, pp. 93-110, Nov. 1996.

[2-syn] S.E. Bou-Ghazale, J.H.L. Hansen, "Synthesis of Stressed Speech from Isolated Neutral Speech Using HMM-Based Models," *ICSLP-96: Inter. Conf. Spoken Lang. Proc.*, vol. 3, pp. 1860-1863, Philadelphia, PA, Oct. 1996.

[3-syn] S. Bou-Ghazale, J.H.L. Hansen, "Improving Recognition and Synthesis of Stressed Speech via Feature Perturbation in a Source Generator Framework," *NATO-ESCA Proc. Inter. Tutorial & Research Workshop on Speech Under Stress,* pp. 45-48, Lisbon, Portugal, Sept. 1995.

[4-syn] S. Bou-Ghazale, J.H.L. Hansen, "Stressed Speech Synthesis with Application to CELP Coding," *EUROSPEECH-95:*

[4] A copy of the workshop paper is contained in `/SUSAS/documentation` under the name `/NATO97-ParserPaper.ps` which has been shown to outperform Entropic's Aligner©phone parsing tool.

PHONEMES USED BY DUKE UNIVERSITY FOR SPEECH TIME ALIGNMENT

DUKE	TIMIT	EXAMPLE	DUKE	TIMIT	EXAMPLE
aa	aa	cot	l	l	led
ae	ae	bat	el	el	bottle
ah	ah	butt	m	m	mom
ao	ao	bought	em	em	'em (syllabic m)
ay	ay	bite	n	n	nun
aw	aw	now	nx	nx	center (nasal flap)
ax	ax,ix, ax-h	the (schwa)	ng	ng, eng	sing
axr	axr	dinner	ow	ow	boat
b	b, bcl	bob	oy	oy	boy
ch	ch	church	p	p, pcl	pop
d	d,dcl	dad	r	r	red
dh	dh	they	s	s	sister
eh	eh	bet	sh	sh	shoe
en	en	button (syllabic n)	t	t, tcl, dx	tot
er	er	bird	th	th	theif
ey	ey	bait	uh	uh	book
f	f	fief	uw	uw, ux	boot, beauty
g	g, gcl	gag	v	v	verve
hh	hh, hv	hay	w	w	wet
ih	ih	bit	y	y	yet
iy	iy	beat	z	z	zoo
jh	jh	judge	zh	zh	measure
k	k, kcl	kick	sil	epi, h#,pau,q	silence

Table 4: Phoneme labeling convention used by Duke University for Speech Time Alignment of SUSAS.

	STRESSFUL SPEECH RECOGNITION RESULTS †												
Condition	*N*	*Sl*	*F*	*So*	*L*	*A*	*C*	*Q*	*C50*	*C70*	*Lom*	*Avg10*	*StDev10*
Stressful, Noise-free	88.3%	60%	65%	48%	50%	20%	68%	75%	63%	63%	63%	57.5%	15.35
Stressful, Noisy	49%	45%	28%	33%	18%	15%	40%	28%	35%	33%	28%	30.3%	9.12

Stressed Speech Styles Key:

N – neutral	So – soft	C – clearly spoken	C50 – Moderate Workload Task Condition
Sl – slow	L – loud	Q – question	C70 – High Workload Task Condition
F – fast	A – angry		Lom – Lombard effect noise condition

Table 5: Recognition performance of neutral and stressed type speech in noise-free and noisy conditions. †Additive white Gaussian noise, SNR = +30dB

ESCA Inter. EUROSPEECH Conf., pp. 455-458, Madrid, Spain, Sept. 1995.

[5-syn] S. Bou-Ghazale, J.H.L. Hansen, "A Source Generator Based Modeling Framework for Synthesis of Speech Under Stress," *ICASSP-95: IEEE Inter. Conf. Acoustics, Speech, Sig. Proc.,* vol. 1, pp. 664-667, Detroit, Michigan, May 1995.

4.3 Stressed Speech Classification

The following papers have considered methods for imparting stress onto neutral input speech.

[1-StCl] B.D. Womack, J.H.L. Hansen, "Classification of Speech Under Stress using Target Driven Features," *Speech Communications,* Special Issue on Speech Under Stress, vol. 20, pp. 131-150, November 1996.

[2-StCl] J.H.L. Hansen, B.D. Womack, "Feature Analysis and Neural Network based Classification of Speech under Stress," *IEEE Trans. on Speech & Audio Proc.,* vol. 4, no. 4, pp. 307-313, July 1996.

[3-StCl] B. Womack, J.H.L. Hansen, "Robust Speech Recognition via Speaker Stress Classification," *ICASSP-96: IEEE Inter. Conf. Acoustics, Speech, Sig. Proc.,* vol. I, pp. 53-56, Atlanta, Georgia, May 1996.

[4-StCl] B. Womack, J.H.L. Hansen, "Robust Stress Independent Speech Recognition using a Neural Network based Stress Classification Algorithm," *EUROSPEECH-95: ESCA Inter. EUROSPEECH Conf.,* pp. 1999-2002, Madrid, Spain, Sept. 1995.

[4-StCl] D. Cairns and J.H.L. Hansen, "Nonlinear Analysis and Detection of Speech Under Stressed Conditions," *Journal Acoust. Soc. of America,* vol. 96, no. 6, pp. 3392-3400, Dec. 1994.

[5-StCl] D. Cairns, J.H.L. Hansen, "Nonlinear Speech Analysis using the Teager Energy Operator with Application to Speech Classification under Stress," *ICSLP-94: Inter. Conf. on Spoken Lang. Proc.,* vol. 3, pp. 1035-1038, Yokohama, Japan, Sept. 1994.

4.4 Stressed Speech Recognition

The following papers have considered methods for improving stressed speech recognition.

[1-recog] J.H.L. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech

Recognition," *Speech Communications,* Special Issue on Speech Under Stress, vol. 20, pp. 151-173, Nov. 1996.

[2-recog] B. Womack, J.H.L. Hansen, "Robust Speech Recognition via Speaker Stress Classification," *ICASSP-96: IEEE Inter. Conf. Acoust., Speech, Sig. Proc.,* vol. I, pp. 53-56, May 1996.

[3-recog] J.H.L. Hansen, S. Bou-Ghazale, "Duration and Spectral Based Stress Token Generation for Keyword Recognition Using Hidden Markov Models," *IEEE Trans. Speech & Audio Proc.,* vol. **3**(5), pp. 415-421, Sept. 1995.

[4-recog] J.H.L. Hansen, M. Clements, "Source Generator Equalization and Enhancement of Spectral Properties for Robust Speech Recognition in Noise and Stress,"*IEEE Trans. Speech & Audio Proc.,* vol. **3**(5), pp. 407-415, Sept. 1995.

[5-recog] S. Bou-Ghazale, J.H.L. Hansen, "Stressed Speech Synthesis with Application to CELP Coding," *EUROSPEECH-95: ESCA Inter. EUROSPEECH Conf.,* pp. 455-458, Madrid, Spain, Sept. 1995.

[6-recog] B. Womack, J.H.L. Hansen, "Robust Stress Independent Speech Recognition using a Neural Network based Stress Classification Algorithm," *EUROSPEECH-95: ESCA Inter. EUROSPEECH Conf.,* pp. 1999-2002, Madrid, Spain, Sept. 1995.

[7-recog] J.H.L. Hansen, D. Cairns, "ICARUS: A Source Generator Based Real-time System for Speech Recognition in Noise, Stress, and Lombard Effect," *Speech Communications,* vol. **16**(4), pp. 391-422, July 1995.

[8-recog] J.H.L. Hansen, "Morphological Constrained Enhancement with Adaptive Cepstral Compensation (MCE-ACC) for Speech Recognition in Noise and Lombard Effect," *IEEE Trans. Speech & Audio Proc.,* SPECIAL ISSUE: Robust Speech Recognition, vol. **2**(4), pp. 598-614, Oct. 1994.

[9-recog] S. Bou-Ghazale and J.H.L. Hansen, "Duration and Spectral Based Stress Token Generation For HMM Speech Recognition under Stress," *ICASSP-94: IEEE Inter. Conf. Acoust., Speech, Sig. Proc.,* vol. 1, pp. 413-416, Adelaide, Australia, April 1994.

[10-recog] J.H.L. Hansen, "Adaptive Source Generator Compensation and Enhancement for Speech Recognition in Noisy Stressful Environments," *ICASSP-93: IEEE Inter. Conf. Acoust., Speech, Sig. Proc.,* vol. II, pp. 95-98, April 1993.

[11-recog] D. Cairns, J.H.L. Hansen, "ICARUS: An Mwave Based Real-time Speech Recognition System in Noise and Lombard Effect," *ICSLP-92: Inter. Conf. on Spoken Lang. Proc.,* vol. II, pp. 703-706, Oct. 1992.

[12-recog] J.H.L. Hansen, O. Bria, "Improved Automatic Speech Recognition in Noise and Lombard Effect," *EURASIP-92, The Sixth European Signal Proc. Conf.,* pp. 403-406, SIGNAL PROCESSING VI: Theory and Applications, Elsevier Publishers, London, U.K., 1992.

[13-recog] J.H.L. Hansen, O. Bria, "Lombard Effect Compensation for Robust Automatic Speech Recognition in Noise," *ICSLP-90: Inter. Conf. on Spoken Lang. Proc.,* pp. 1125-1128, Kobe, Japan, Nov. 1990.

[14-recog] J.H.L. Hansen, M.A. Clements, "Stress Compensation and Noise Reduction Algorithms for Robust Speech Recognition," *ICASSP-89: IEEE Inter. Conf. Acoust., Speech, Sig. Proc.,* pp. 266-269, Glasgow, Scotland, May 1989.

SUMMARY & CONCLUSIONS

In this paper, we have discussed the problem of speech under stress, which includes aspects of emotion, task workload stress, and *Lombard* effect. A database entitled SUSAS for *Speech Under Simulated and Actual Stress* was presented, including details of how various domains of speech data were collected. The intent of this speech corpus is to assist researchers in the analysis, modeling, and development of speech algorithms which address the issues of stress, emotion, and Lombard effect. A number of previous studies were listed in the areas of analysis, classification, synthesis, and recognition of speech under stress. While most of these studies focused on simulated stressed speech data, it is important to understand that the levels of stress experienced in actual high-stress, noisy, or adverse environments may be more extreme than those under simulated conditions. As such, some analysis has been conducted on actual speech under stress. We hope this data will prove to be useful for other researchers developing robust algorithms to address speaker variability under stress.

ACKNOWLEDGMENT

I wish to extend thanks to a number of individuals who provided assistance during the collection and organization of SUSAS. First, I would like to thank the student volunteers from the Georgia Tech signal processing laboratory, who allowed me to record their speech during roller-coaster ride tasks. I would also like to acknowledge Mark Clements for his support during my graduate studies at Georgia Tech, and for introducing me to problems in speech enhancement and recognition under stress. I would also like to thank a number of my students, especially Bryan Pellom and Ruhi Sarikaya who assisted in re-digitizing large portions of the actual speech under stress data, as well as labeling simulated and actual stress, Doug Cairns for establishing real-time speech recognition scores under stress[5, 6], Brian Womack for evaluations on stress classification using SUSAS, and Sahar Bou-Ghazale who has helped organize and maintain the NATO speech under stress Web page during 1996-97.

WEB INFORMATION

The Duke Robust Speech Processing Lab is maintaining the NATO Speech Under Stress web page at the following URL. Please check this location for updates on studies which have used the SUSAS stressed speech data.

http://www.ee.duke.edu/Research/Speech/stress.html

FUTURE RELEASE of SUSAS_2.0

This release of SUSAS (1.4) contains label files and speech for simulated and actual stress. Because of the wide variations in recording and speaker characteristics (some files were amplitude clipped, etc.) from the Actual portion of SUSAS, some files originally included on the tapes were not included in the main releases of SUSAS_1.0 (up to 1.4). We have recently completed a re-digitization all tapes from the Actual portion and are presently in the process of building a more extensive image of the Actual sub-directory [more training data, label files hand corrected, etc.]. We will make this available as SUSAS_2.0 during the fall of 1998 for those interested in further studies on Speech Under Stress. See the NATO Web page above for more information.

References

[1] Z.S. Bond, T.J. Moore, "A note on Loud and Lombard Speech," *1990 ICSLP,* pp.969-972.

[2] S.E. Bou-Ghazale, "Duration and Spectral based Stress Token Generation for Keyword Recognition using Hidden Markov Models," M.S. Thesis, Robust Speech Processing Lab, Dept. of Electrical Engineering, Duke Univ., June 1993.

[3] S.E. Bou-Ghazale, J.H.L. Hansen, "Duration and Spectral Based Stress Token Generation For HMM Speech Recognition under Stress," *IEEE 1994 ICASSP,* pp. 413-416.

[4] O. Bria, "Improved Automatic Speech Recognition Under Lombard Effect," M.S. Thesis, Robust Speech Processing Lab, Dept. of Electrical Engineering, Duke Univ., April 1991.

[5] D.A. Cairns, "Real-time Speech Recognition under Lombard Effect and in Noise," M.S. Thesis, Robust Speech Processing Lab, Dept. of Electrical Engineering, Duke Univ., April 1991.

[6] D.A. Cairns and J.H.L. Hansen, "ICARUS: An Mwave Based Real-time Speech Recognition System in Noise and Lombard Effect," *ICSLP-92, Inter. Conf. Spoken Lang. Proc.,* pp. 703-706, Alberta, Canada, October 1992.

[7] D.A. Cairns, J.H.L. Hansen, "Nonlinear Analysis and Detection of Speech Under Stressed Conditions," *J. of Acoust. Soc. of America,* vol.96, no.6, pp. 3392-3400, Dec. 1994.

[8] Y. Chen, "Cepstral Domain Talker Stress Compensation for Robust Speech Recognition," *IEEE Trans. on ASSP,* pp.433-439, April 1988.

[9] G.J. Clary, J.H.L. Hansen, "A Novel Speech Recognizer for Keyword Spotting," *1992 ICSLP,* pp.13-16.

[10] J.K. Darby, *Speech Evaluation in Psychiatry,* Grune & Stratton, New York, New York, 1981.

[11] M. Flack, "Flying Stress," London: Medical Research Committee, 1918.

[12] D.J. Folds, J.M. Gerth, W.R. Engelman, " Enhancement of Human Performance in Manual Target Acquisition and Tracking, " Final Technical Report USAFASM-TR-86-18, USAF School of Aerospace Medicine, Brooks AFB, TX, 1986.

[13] D.J. Folds, " Response Organization and Time-Sharing in Dual-Task Performance," Ph.D. dissertation, School of Psychology, Georgia Institute of Technology, Atlanta, May 1987.

[14] M.B. Gardner, "Effect of Noise System Gain, and Assigned Task on Talking Levels in Loudspeaker Communication," *J. Acoust. Soc. Am.,* **40**:955- 965, 1966.

[15] C.N. Hanley, D.G. Harvey, "Quantifying the Lombard Effect," *J. of Hearing & Speech Disorders*, **30**:274-7, Aug. 1965.

[16] J.H.L. Hansen, "Morphological Constrained Enhancement with Adaptive Cepstral Compensation (MCE-ACC) for Speech Recognition in Noise and Lombard Effect," *IEEE Trans. Speech & Audio,* SPECIAL ISSUE: Robust Speech Recognition, **2**(4):598-614, 1994.

[17] J.H.L. Hansen, "Adaptive Source Generator Compensation and Enhancement for Speech Recognition in Noisy Stressful Environments," *IEEE 1993 ICASSP,* pp. 95-98.

[18] J.H.L. Hansen, "Evaluation of Acoustic Correlates of Speech Under Stress for Robust Speech Recognition." *IEEE Proc. 15th Bioengineering Conf.,* pp. 31-32, Boston, Mass., March 1989.

[19] 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., Atlanta, GA, 428 pgs., July 1988.

[20] J.H.L. Hansen, O. Bria, "Improved Automatic Speech Recognition in Noise and Lombard Effect," *EURASIP-92.* In *Signal Processing VI: Theories and Applications,* Elsevier Publishers, New York, NY, pp. 403-406, 1992.

[21] J.H.L. Hansen, O.N. Bria, "Lombard Effect Compensation for Robust Automatic Speech Recognition in Noise," *Proc. Inter. Conf. Spoken Lang. Proc.,* pp. 1125-8, Kobe, Japan, Nov. 1990.

[22] J.H.L. Hansen, D.A. Cairns, "ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments," *Speech Communications,* **16**:391-422, July 1995.

[23] J.H.L. Hansen, M.A. Clements, "Stress Compensation and Noise Reduction Algorithms for Robust Speech Recognition," *IEEE 1989 ICASSP,* pp. 266-9.

[24] J.H.L. Hansen, M.A. Clements, "Evaluation of Speech under Stress and Emotional Conditions," *Proc. Acoust. Soc. Am.,* H15, **82**(Fall Sup.):S17, Nov. 1987.

[25] J.H.L. Hansen, B.D. Womack, "Feature Analysis and Neural Network-Based Classification of Speech under Stress," *IEEE Trans. on Speech & Audio Processing,* vol. 4, no. 4, pp. 307-313, July 1996.

[26] B.A. Hanson, T. Applebaum, "Robust Speaker-Independent Word Recognition Using Instantaneous, Dynamic and Acceleration Features: Experiments with Lombard and Noisy Speech," *IEEE 1990 ICASSP,* 857–60.

[27] M.H.L. Hecker, K.N. Stevens, G. von Bismarck, C.E. Williams, "Manifestations of Task-Induced Stress in the Acoustic Speech Signal," *J. Acoust. Soc. Am.,* **44**(4):993-1001, 1968.

[28] H. Hermansky, N. Morgan, H.G. Hirsch, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing," *IEEE 1993 ICASSP,* pp.83-86.

[29] J.W. Hicks, H. Hollien, "The Reflection of Stress in Voice-1: Understanding the Basic Correlates," *1981 Carnahan Conf. on Crime Countermeasures,* 189-195, 1981.

[30] M.J. Hunt, C. Lefebvre, "A comparison of several acoustic representations for speech recognition with degraded and undegraded speech," *IEEE 1989 ICASSP,* pp.262-5.

[31] H.R. Jex, " A Proposed Set of Standardized Sub-Critical Tasks For Tracking Workload Calibration, " in N. Moray, *Mental Workload: Its Theory and Measurement*, New York: Plenum Press, pp. 179-188, 1979.

[32] B.H. Juang, "Speech Recognition in Adverse Environments," *Computer, Speech & Lang.,* pp.275-94, 1991.

[33] J.C. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers," *J. Acoust. Soc. Am.,* (1):510-24, 1993.

[34] I. Kuroda, O. Fujiwara, N. Okamura, N. Utsuki, "Method for Determining Pilot Stress Through Analysis of Voice Communication," *Aviation, Space, & Env. Med.,* **5**:528-533, 1976.

[35] P. Lieberman, S. Michaels, "Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech," *J. Acoust. Soc. Am.,* **34**(7):922-7, 1962.

[36] R.P. Lippmann, M. Mack, D. Paul, " Multi-Style Training for Robust Speech Recognition Under Stress, " *Proc. of the Acoustical Society of America*, 110th Meeting, QQ10, May, 1986.

[37] R.P. Lippmann, E.A. Martin, D.B. Paul, "Multi-Style Training for Robust Isolated-Word Speech Recognition," *IEEE 1987 ICASSP,* pp.705-8.

[38] S. Lively, D. Pisoni, W. van Summers, R. Bernacki, "Effects of cognitive workload on speech production," *J. Acoust. Soc. Am.,* **93**(5) 2962-73, 1993.

[39] F.H. Liu, A. Acero, R.M. Stern, "Efficient joint compensation of speech for the effects of additive noise and linear filtering," *IEEE 1992 ICASSP,* pp. 257–60.

[40] E. Lombard, " Le Signe de l'Elevation de la Voix, " *Ann. Maladies Oreille, Larynx, Nez, Pharynx,* **37**:101-19, 1911.

[41] D. Mansour, B.H. Juang, "A Family of Distortion measures based upon projection operation for robust speech recognition," *IEEE Trans. ASSP,* **37**:1659–71, 1988.

[42] D.B. Paul, "A Speaker-Stress Resistant HMM Isolated Word Recognizer," *IEEE 1987 ICASSP,* pp.713–6.

[43] D.B. Paul, C.J. Weinstein, R.P. Lippman, Y. Chen, "Robust HMM-Based Techniques for Recognition of Speech Produced Under Stress and in Noise," *Proc. Speechtech-86 Conf.*, pp. 241-249, April, 1986.

[44] B.L. Pellom, J.H.L. Hansen, "Automatic Segmentation and Labeling of Speech Recorded in Unknown Noisy Channel Environments," *ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels,* pp. 167-170, Pont-a-Mousson, France, April 1997.

[45] B.L. Pellom, J.H.L. Hansen, "Automatic Segmentation of Speech Recorded in Unknown Noisy Channel Characteristics," accepted to *Speech Communication,* Feb. 1998. To appear Fall 1998.

[46] D.B. Pisoni, et. al, "Some Acoustic-Phonetic Correlates of Speech Produced in Noise," *IEEE 1985 ICASSP,* 41.10.1-4.

[47] P.K. Rajasekaran, G.R. Doddington, J.W. Picone, "Recognition of Speech Under Stress & Noise," *IEEE 1986 ICASSP,* pp.733–6.

[48] B.J. Stanton, L.H. Jamieson, G.D. Allen, "Acoustic-Phonetic Analysis of Loud and Lombard Speech in Simulated Cockpit Conditions," *IEEE 1988 ICASSP,* pp.331–4.

[49] B.J. Stanton, L.H. Jamieson, G.D. Allen, "Robust Recognition of Loud and Lombard Speech in the Fighter Cockpit Environment," *IEEE 1989 ICASSP,* pp.675–8.

[50] L.A. Streeter, N.H. Macdonald, W. Apple, R.M. Krauss, K.M. Galotti, "Acoustic and Perceptual Indicators of Emotional Stress," *J. Acoust. So. Am.* **73**(4) 1354-1360, 1983.

[51] B.D. Womack, "Classification and Recognition of Speech under Perceptual Stress using Neural Networks and N-D HMMs," Ph.D. Thesis, Robust Speech Processing Lab, Dept. of Electrical Engineering, Duke Univ., Dec. 1996.

[52] C.E. Williams, K.N. Stevens, "On Determining the Emotional State of Pilots During Flight: An Exploratory Study," *Aerospace Medicine,* **40** 1369-1372, 1969.

[53] C.E. Williams, K.N. Stevens, "Emotions and Speech: Some Acoustic Correlates," *J. Acoust. Soc. Am.,* **52**(4):1238-50, 1972.