'Houston, We have a solution' : Using NASA Apollo Program to advance Speech and Language Processing Technology

Abhijeet Sangwan[1], *Lakshmish Kaushik*[1], *Chengzhu Yu*[1], *John H. L. Hansen*[1], *and Douglas W. Oard*[2]

[1]Center for Robust Speech Systems (CRSS), The University of Texas at Dallas,
Richardson, TX, U.S.A.
[2] College of Information Studies and UMIACS, University of Maryland,
College Park, MD, U.S.A.

{abhijeet.sangwan,lakshmish.kaushik,cxy110530,john.hansen}@utdallas.edu, oard@umd.edu

Abstract

NASA's Apollo program stands as one of mankind's greatest achievements in the 20th century. During a span of 4 years (from 1968 to 1972), a total of 9 lunar missions were launched and 12 astronauts walked on the surface of the moon. It was one the most complex operations executed from scientific, technological and operational perspectives. In this paper, we describe our recent efforts in gathering and organizing the Apollo program data. It is important to note that the audio content captured during the 7-10 day missions represent the coordinated efforts of hundreds of individuals within NASA Mission Control, resulting in well over 100k hours of data for the entire program. It is our intention to make the material stemming from this effort available to the research community to further research advancements in speech and language processing. Particularly, we describe the speech and text aspects of the Apollo data while pointing out its applicability to several classical speech processing and natural language processing problems such as audio processing, speech and speaker recognition, information retrieval, document linking and a range of other processing tasks which enable knowledge search, retrieval, and understanding.. We also highlight some of the outstanding opportunities and challenges associated with this dataset. Finally, we also present initial results for speech recognition, document linking, and audio processing systems.

Index Terms: NASA Apollo program, Speech Processing, Natural Language Processing, Audio Processing

1. Introduction

NASA's Apollo program stands as one of mankind's greatest achievements in the 20th century. During a span of 4 years (from 1968 to 1972), a total of 9 lunar missions were launched and 12 astronauts walked on the surface of the moon, with as many as 400,000 personnel supporting the program in various capacities. It was arguably one the most complex operations executed from scientific, technological and operational perspectives. Furthermore, these operations were meticulously recorded, documented, annotated and discussed (*e.g.* interviews, oral histories, press conferences *etc.*). For example, the mission produced over 10,000 technical documents and (considering all 15 missions flown with the "Apollo system") over 200 days of multi-channel audio data, in addition to photographs, videos, telemetry, and other forms of data. The audio and text material captured the complex interaction between astronauts, mission control, scientists, engineers and others involved in the program. Apollo data is unique in that it is perhaps one of the few such data sources that is available in the public domain, thereby making the study of large scale complex events feasible. For example, more recent historical events such as the Hurricane Katrina disaster, 9/11 Terrorist Attacks or the Fukushima Daiichi nuclear reactor meltdown bear resemblance to the Apollo missions in terms of the number of personnel involved, criticality of the operation, complexity of the undertaking, and the degree of inter-communication required. However, access to these data sources for research and scientific study may be difficult, if not impossible.

The Apollo data offers many opportunities to pursue advancements in Human Language Technology (HLT). It offers interesting possibilities for content linking, where audio-to-text as well as audio-to-audio links can be established based on contextual similarity and relevance. Integrating these time-synchronized sources with topic-alignable external resources (including press conference and change-of-shift briefing audio, but also the thousands of available scanned documents) is an exceptionally challenging information integration task. A content linking capability can spawn powerful interfaces which organize the relevant knowledge in a manner that provides easy and convenient access. It is worth mentioning that such access would be equally meaningful for amateur enthusiasts and professional investigators. For example, a person reviewing oral history videos could automatically gain access to relevant portions of the mission data, or a technical investigator could obtain access to the design document of engineering objects being discussed during the mission.

The interesting opportunities associated with the Apollo data also bring forth some challenges. In the Apollo dataset, the onboard audio recordings and space-to-ground communications are among the most challenging publicly available large-scale collections of time-critical and mission-critical audio. The audio material was collected in the presence of highly variable background noise and channel conditions, posing significant real-world challenges to existing speech processing algorithms/techniques. In summary, the audio data presents classical speech and audio processing problems, albeit, in a real-world setting.

It is our intention to make this dataset available to the research community. We intend to create awareness of this dataset, and the merits, challenges and opportunities it presents. In the remainder of this document, we describe the dataset in further detail. We also describe the research problems that we are currently pursuing with this dataset. Finally, we present initial results for various speech and language processing tasks.

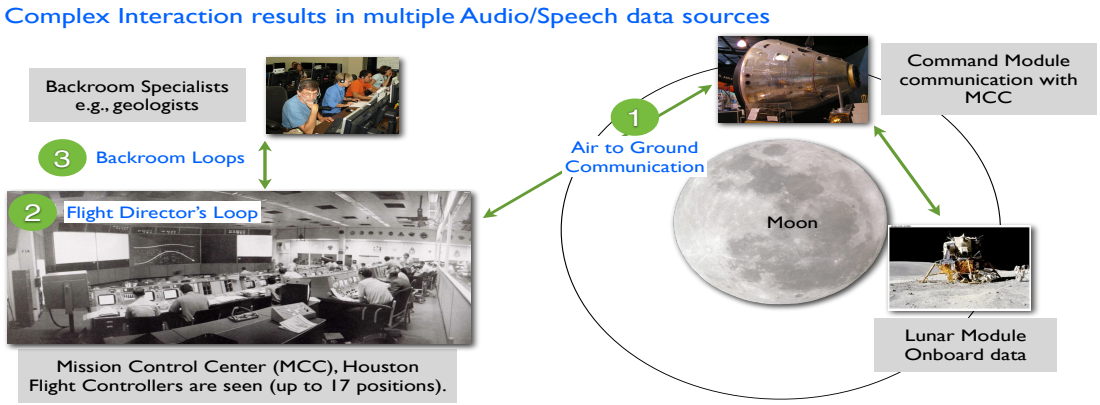

Figure 1: Apollo audio data reflects the complex interaction between mission personnel. Astronauts (from command and lunar modules) communicated with Mission Control Center (MCC). Flight controllers within MCC communicated with the flight director. Each flight controller communicated with his backroom support which consisted of engineers, scientists, *etc.*. An interactive layout of the MCC can be accessed here: `http://1.usa.gov/WY989U`

2. Apollo Missions and Data Sources

2.1. Audio Data

The audio data collection can broadly be divided into two datasets: (i) data captured during the mission and (ii) data captured after and before the mission. The data captured during the mission was due to the complex interaction between the core mission personnel, *i.e.*, astronauts, flight controllers, and backroom specialists (as shown in Fig. 1). The 9 lunar missions lasted between 6 and nearly 13 days (*e.g.*, Apollo 11 lasted 8 days 3 hours 18 minutes and 35 seconds). In what follows, these audio data sources are explained in more detail.

2.1.1. Onboard audio

The Apollo program included two spacecraft: the Lunar Module (LM) to land on the moon and the Command Module (CM), which remained in lunar orbit [1] (see Fig. 1). Each contained a combined voice and data recorder. The CM recorder used far-field microphones to record crew activity when radio communication was not possible (e.g., behind the Moon); the tape was replayed at high speed over a low-margin radio link when radio communication was reestablished. The LM recorder could not be rewound in flight, so its ten-hour capacity was selectively used during critical parts of the mission and the tape was returned to Earth.

2.1.2. Mission Control Center (MCC) audio

As many as 17 positions in the MCC were staffed at various times during the mission, and these flight controllers communicated among themselves and with "back room" specialists over dedicated intercom circuits (loops) (see. Fig. 1). Much of the complexity of managing a mission occurred on these loops. Speakers used close talking microphones.[2]

2.1.3. Space-to-ground communication

The 9 Apollo lunar missions lasted between 6 and nearly 13 days. Because of the trajectory, communication with the spacecraft was possible for about 90% of this time. These recordings exhibit highly variable channel characteristics because several different receiving stations and relay facilities were used [3]. Speakers used close-talking microphones. Two versions are available, one with superimposed public affairs commentary.

Additionally, audio data is also available from pre- and post-mission sources such as:

2.1.4. Press conferences

Each mission included pre-mission and post-mission press conferences with the astronauts, and change-of-shift briefings by flight controllers. Many of these events include responses to audience questions, and they were generally recorded with far-field microphones.[4]

2.1.5. Debriefs

Following each mission, the astronauts conducted a set of structured discussions in which specific aspects of the mission were reviewed, most recorded with far-field microphones.

2.1.6. Interviews

The NASA History Division has conducted interviews with about 270 Apollo participants, including astronauts, controllers, engineers, and managers, many with near-field microphones.[5]

About 30% of this audio is presently available online; the remainder is available only on physical media from the NASA Johnson Space Center (JSC) Media Resource Center [20] or (for the oral history interviews) the JSC History Office collection at the University of Houston, Clear Lake.

2.2. Text Data

In addition to the large audio collection, a vast amount of text is also available. Some of the prominent sources are:

[1] Sample onboard audio: `http://1.usa.gov/WPFl2j`
[2] Sample MCC audio: `http://bit.ly/XtxDGR`
[3] Sample space-to-ground: `http://1.usa.gov/13gMEno`
[4] Sample press conference: `http://bit.ly/YJaqA4`
[5] Sample oral history interview: `http://cs.pn/13gMTlN`

2.2.1. *Transcripts*

Space-to-ground and onboard audio were transcribed for engineering analysis, and PAO commentary was transcribed for press releases; all (except some onboard transcripts) have been processed using Optical Character Recognition (OCR), corrected, and extensively annotated. [6]

2.2.2. *Debriefs*

Debriefs were also transcribed and are generally available in scanned but uncorrected form. MCC loop audio was not generally transcribed. Digital transcripts are available for oral history interviews, but few press conferences seem to have transcripts available.

2.2.3. *Technical reports*

More than 3,000 Apollo Program technical reports have been scanned, and another 10,000 could be scanned on request. These documents address a broad range of issues, including the design of, procedures for, management of, training for, and experience with Apollo spacecraft, launch vehicles, and ground facilities.

2.2.4. *Digital Books/Manuscripts*

Several dozen digital books and manuscripts that address the Apollo program are available online from the NASA History Division.

While the goal of capturing and organizing audio and document content from the entire Apollo program is an attractive mission, a major challenge exists before the entire effort can be fully realized. This represents extracting the audio from the 30-track synchronized recordings from NASA. As it turns out, the 30-track record/playback system used during the 12 year span does not presently exist. Only a 2-track playback system is available, and since the time track must be captured during each play-back, the digitizing process to extract this audio corpus will require re-digitizing each tape 29 times (since each mission might last 192 hours with 30tracks operating for each hour, this would require a total of 5568 hours of digitizing for each Apollo mission). Presently, our group is working to reconstruct a high-end playback system which would streamline this process, and allow collaboration with NASA to expedite the digitizing process. Additional challenges include ensuring the highest playback quality as well as safeguarding the unique tapes.

3. Research Opportunities and Challenges

The sheer volume and complexity of the NASA Apollo data and the underlying operation provides many research opportunities and challenges for audio, speech and language processing. One fruitful area of research is content linking, which automatically binds text, video, or still image documents to audio material (and vice-versa) in a context-sensitive manner. While text-to-text content linking has been well researched, topic-based linking from spoken content has not yet been well investigated [1, 2, 3, 4]. Using the Apollo data, a number of relevant text-to-audio linking problems can be defined. Another interesting problem along similar lines is to synchronize two audio sources that share the same timeline (in absence of the time code). Here,

[6] Example transcript: `http://1.usa.gov/Xtymb3`

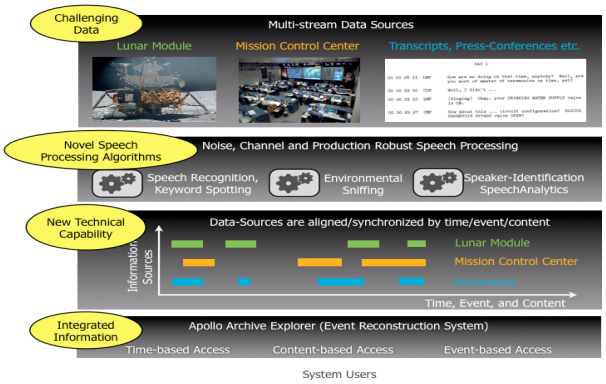

Figure 2: Apollo data provides the opportunity to develop new information integration capability within an event reconstruction system using novel robust speech and language processing algorithms.

several variants of this task exist, but the most challenging is to align one side of a conversation (e.g., words spoken by astronauts as recorded onboard the spaceship) with a two-sided recording of the same conversation (e.g., recordings of the radio circuit over which some but not all of those words were spoken) where recordings have markedly different channel characteristics (not to mention the relative time location of each speaker). Furthermore, selective integration of the space-to-ground audio with more than a dozen internal Mission Control Center (MCC) audio intercom "loops" is a complex and challenging audio integration task. Along similar lines, metadata extraction and information retrieval are other possible areas of research, which has been extensively explored for many domains such as broadcast news (BN), e-learning, parliamentary proceedings, sports (tennis, soccer, basketball etc.), movies, TV-shows, robotics, and security [5, 15, 7, 12].

The dataset also offers tremendous potential for traditional speech processing technology such as robust speech and speaker recognition. In the context of Apollo, this is a hard problem given that the audio has been observed to contain several artifacts such as (i) variable additive noise, (ii) variable channel characteristics, (iii) reverberation, and (iv) variable bandwidth accompanied with speech production issues (such as stress), and speech capture issues (such as astronaut speech captured while walking on the moon in spacesuits). Hence, this dataset would be different from traditional datasets (where all or at least a large proportion of the dataset is telephone speech). Additionally, the ability to use the dataset to support upstream NLP (natural language processing) tasks (and other interesting applications) can also foster new collaborative research effort where the knowledge extracted by speech and speaker recognition systems is employed to drive other systems.

In speech processing, most applications are focused on audio stream processing within a short context (i.e., several seconds to perhaps minutes or at most hours). Another domain of interest is speaker state analysis over several days, leading to the completion of a major task such as walking on the moon, along with a safe return. A number of studies have considered detection of speech under stress [18, 19, 20] and some over a short task period [23], or recognition of speech under stress [21, 22], but no studies to date have considered speaker state analysis of individuals over an extended critical period such as an Apollo mission. The diversity and variability of speaker state for astronauts over a 7-10 day mission offer a unique opportunity in monitoring individuals through voice communications.

Another possible area of research could be acoustic environment detection. Researchers have pursued environment estimation with an intent of generating relevant acoustic contextual knowledge to adjust/tune speech systems in order to deliver better performance [8, 9, 14]. Additionally, research has also been conducted on scene recognition system which use audio and other information sources to estimate the environment [10, 11]. Automatic estimation of background information in long-term audio has also been used to generate daily summaries and analysis for Life-Logging applications[13]. Finally, while stress detection represents an important tracking of speaker traits, the general problem of speaker variability for speaker identification (SID) represents a new challenging area for the current data context. Speaker recognition of both astronauts and scientists/engineers/mission specialists over extended periods with corresponding task stress, fatigue, zero gravity, channel/noise, and a range of other factors represents a new paradigm not ever seen in the most popular speaker ID evaluations (i.e., NIST SRE efforts). The resulting tagging of speakers, with both location, context, and potentially speaker state would offer a unique opportunity to explore new advancements for SID in a truly diverse range of speaker variability.

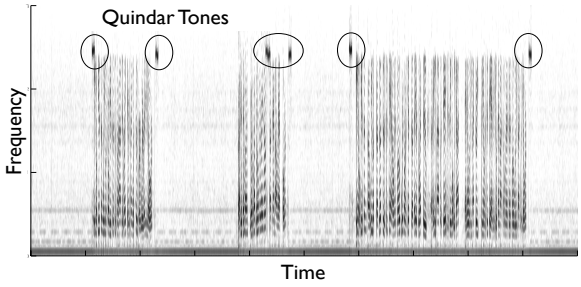

Figure 3: Quindar Tones were a means of in-band signaling for remote transmitted keying and release when transmitting from Earth.

4. Results

In this section, we present initial results on the Apollo dataset for various tasks.

4.1. Document Linking

To establish a baseline for content linking, we have built a simple system for linking a passage from a mission transcript with corresponding descriptions of the same event from manualy transcribed post-mission oral gistory interviews [17]. When evaluated using Mean Reciprocal Rank (which awards partial credit for placing the proper interview passage second or third in a ranked list) for cases in which there was a known interview passage to be found, values of about 0.5 were obtained. This equates to an expected rank of 2, where random performance would yield a rank below 4,000. For practical reasons, these initial experiments were conducted on an uncorrected OCR version of the mission transcript, and only a small range of fixed context widths were tried. We therefore believe that even stronger baselines can be constructed.

4.2. Acoustic Signal Processing

The Apollo missions used quindar tones for in-band transmitter keying. The quindar tones can be heard before and after the Capsule Communicator (CapCom) communicates with the astronauts. Specifically, two types of Quindar tones were used, "intro tone" and "outro tone". Fig. 3 shows the time-frequency signature of the quindar tones. The higher frequency "intro tone" at 2525Hz was used to activate the ground-to-space transmissions. The slightly lower frequency "outro tone" at 2475 was used to turn off the ground-to-space transmissions. Both "intro tone" and "outro tone" were 250ms long. Qunidar tones are of interest because of their significance for several audio analysis tasks such as synchronizing audios from different audio tracks or recovering a true sampling rate that was corrupted when played back in a repaired 30-track Soundscriber machine from 1960. Therefore, we developed a Quindar tone detection tool by modifying traditional single tone detection algorithm to be robust to distortions caused by channel noise as well as pitch shifting [16]. Based on an initial evaluation test on the Apollo 11 mission audio, approximate of 90% of the Quindar tones could be detected ar about a 10% false alarm rate.

4.3. Speech Recognition

We setup an initial ASR probe experiment with Apollo 11 data. We gathered text data from the sources mentioned in Sec. 2.2 (excluding Apollo 11) for the language model. We had to run modern OCR systems on original documents as the previously OCRd documents had significant number of errors. This process generated better quality text which was then used to train a 3-gram language model of 38k words. Additionally, we prepared 11 hours of audio data with corresponding ground truth transcripts. Since the time alignments in the transcripts are known to be problematic, we ran forced alignment to eliminate possibly incorrect (or misaligned) transcripts. It was observed that about 75% of the transcripts (8 hours) failed forced alignment due to various issues such as (i) OCR errors (ii) background noise and/or (iii) unfaithful transcripts. The remaining 3 hours of data was split equally into adapt and evaluation sets. A conversational telephone acoustic model (trained on a mixture of switchboard and fisher datasets) was used as baseline, and using the adapt set, MLLR followed by MAP adaptation was executed. We obtained a word error rate (WER) of 92% and 77% for the baseline and adapted systems, respectively. Our experimental results reveal the difficulty of this task. In order to achieve an effective speech recognizer, high quality transcripts are necessary and generating this for Apollo 11 is non-trivial (inspite of transcription availability). We are currently exploring solutions that leverage other capabilities such as quindar tone detection to extract precision aligned transcripts.

5. Conclusion

The Apollo dataset contains enormous opportunities and challenges. In this paper, we demonstrated how the Apollo dataset can foster research in document linking (both text-to-audio and audio-to-audio). We have also shown that the dataset may be interesting to speech and speaker recognition community, as it poses challenging conditions such as channel variability, time-varying background noise, and speech production variability. We are confident that we have merely scratched the surface in terms of identifying the various novel problems that can be elegantly solved using the Apollo dataset. More importantly, we also wish to point out that this dataset is like no other in terms of complexity, scope, significance, potential and availability. Currently, we are working towards gathering, organizing and publishing this material with an intention of making it available to the research community.

6. References

[1] J. Mayfield, D. Alexander, B. Dorr, J. Eisner, T. Elsayed, T. Finin, C. Fink, M. Freedman, N. Garera, P. McNamee, S. Mohammad, D. W. Oard, C. Piatko, A. Sayeed, Z. Syed, R. Weischedel, T. Xu and D. Yarowsky, Cross-Document Coreference Resolution: A Key Technology for Learning by Reading, in AAAI Spring Symposium on Learning by Reading and Learning to Read, Stanford, CA, 2009.

[2] A. Sayeed, T. Elsayed, N. Garera, D. Alexander, T. Xu, D. W. Oard, D. Yarowsky and C. Piatko, Arabic Cross-Document Coreference Resolution, Annual Conf. Assoc. for Computational Linguistics/Inter. Joint Conference on Natural Language Processing, pp. 357-360, Singapore, 2009.

[3] P. McNamee, J. Mayfield, D. Lawrie, D. W. Oard and D. Doermann, Cross-Language Entity Linking, Fifth Inter. Joint Conference on Natural Language Processing, Chaing Mai, Thailand, 2011.

[4] T. Xu and D. W. Oard, "Wikipedia-Based Topic Clustering for Microblogs," in 74th Annual Conference of the American Society for Information Science and Technology, New Orleans, LA, 2011.

[5] M. Naphade, J. R. Smith, J. Tesic, S. F. Chang, W. Hsu, L. Kennedy, A. Hauptmann and J. Curtis, Large-Scale Concept Ontology for Multimedia, IEEE Multimedia Magazine, Vol. 13, No. 3, pp. 86-91, 2006.

[6] F. De Jong, R. Ordelman and M. Huijbregts Automated speech and audio analysis for semantic access to multimedia, Lecture Notes in Computer Science, Vol. 4306, pp. 226-240, 2006,

[7] M. Cristani, M. Bicego and V. Murino, Audio-visual event recognition in surveillance video sequences, IEEE Transactions on Multimedia, Vol. 9, No. 2, pp. 257-267, 2007.

[8] N. Krishnamurthy and J. H. L. Hansen, Babble noise:modeling, analysis, and applications, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, Num. 7, pp. 1394-1407, 2009.

[9] H. Boril, N. Krishnamurthy and J.H.L. Hansen, Online Noise and Lombard Effect Compensation for In-Vehicle Automatic Speech Recognition, 4th Biennial Workshop on DSP for In-Vehicle Systems and Safety, Dallas, TX, USA, 2009.

[10] S. Chu, S. Narayanan, C.-C. J. Kuo and M. J. Mataric, Where am I? Scene recognition for mobile robots using audio features, IEEE International Conference on Multimedia and Expo (ICME), pp. 885-888, Toronto, Canada, 2006.

[11] Wang, DeLiang, and Guy J. Brown, eds. Computational auditory scene analysis: Principles, algorithms, and applications. IEEE Press, 2006.

[12] J. H. L. Hansen, R. Huang, B. Zhou, M. Seadle, J. R. Deller Jr., A. R. Gurijala and P. Angkititrakul, SpeechFind: Advances in Spoken Document Retrieval for a National Gallery of the Spoken Word, IEEE Trans. Speech and Audio Processing, Special Issue on Data Mining, vol. 13, no. 5, pp. 712 - 730, 2005

[13] A. Ziaei, A. Sangwan, John Hansen, "Prof-Life-Log: Personal Interaction Analysis on Naturalistic Audio Streams. Submitted to ICASSP2013.

[14] M. Akbacak, J. H. L. Hansen, ENVIRONMENTAL SNIFFING: Robust Digit Recognition for an In-Vehicle Environment, INTERSPEECH-2003/Eurospeech-2003, pp. 2177-2180, Geneva, Switzerland, 2003.

[15] H. Boril, A. Sangwan, T. Hassan and J. H. L. Hansen, Automatic Excitement-Level Detection for Sports Highlights Generation, 11th Annual Conference on the International Speech Communication Association (Interspeech-2010), Makuhari, Japan, 2010.

[16] Y. T. Chan, Q. Ma, H. C. So, and R. Inkol, Evaluation of Various FFT Methods for Single Tone Detection and Frequency Estimation, IEEE Canadian Conference on Electrical and Computer Engineering, Vol. 1, May 1997, pp. 211-214.

[17] D. Oard et. al., "Linking Transcribed Conversational Speech", submitted to The 36^{th} Annual ACM SIGIR Conference, Dublin, Ireland, August 2013.

[18] G. Zhou, J.H.L. Hansen, and J.F. Kaiser, "Nonlinear Feature Based Classification of Speech under Stress," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 2, pp. 201-216, March 2001.

[19] J.H.L. Hansen, B. Womack, "Feature Analysis and Neural Network based Classification of Speech under Stress," IEEE Transactions on Speech and Audio Processing, vol. 4, no. 4, pp. 307-313, July 1996.

[20] D. Cairns, J.H.L. Hansen, "Nonlinear Analysis and Detection of Speech Under Stressed Conditions," The Journal of the Acoustical Society of America, vol. 96, no. 6, pp. 3392-3400, December 1994.

[21] 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(2), pp. 151-170, November 1996.

[22] J.H.L. Hansen, M. Clements, "Source Generator Equalization and Enhancement of Spectral Properties for Robust Speech Recognition in Noise and Stress," IEEE Transactions on Speech and Audio Processing., vol. 3, no. 5, pp. 407-415, Sept. 1995.

[23] J.H.L. Hansen, E. Ruzanski, H. Boril, J.. Meyerhoff, "TEO-based Speaker Stress Assessment using Hybrid Classification and Tracking Schemes," Inter. Journal Speech Technology, vol. 15, issue 3, pp. 295-311, Sept. 2012