



‘Houston, We have a solution’: A Case Study of the Analysis of Astronaut Speech during NASA Apollo 11 for Long-Term Speaker Modeling

Chengzhu Yu¹, John H. L. Hansen¹ and Douglas W. Oard²

¹Center for Robust Speech Systems (CRSS)
The University of Texas at Dallas, Richardson, TX, U.S.A.

²College of Information Studies and UMIACS
University of Maryland, College Park, MD, U.S.A.
{chengzhu.yu, john.hansen}@utdallas.edu, oard@umd.edu

Abstract

Speech and language processing technology has the potential of playing an important role in future deep space missions. To be able to replicate the success of speech technologies from ground to space, it is important to understand how astronaut’s speech production mechanism changes when they are in space. In this study, we investigate the variations of astronaut’s voice characteristic during NASA Apollo 11 mission. While the focus is constrained to analysis of the three astronauts voices who participated in the Apollo 11 mission, it is the first step towards our long term objective of automating large components of space missions with speech and language technology. The result of this study is also significant from an historical point of view as it provides a new perspective of understanding the key moment of human history - landing a man on the moon, as well as employed for future advancement in speech and language technology in “non-neutral” conditions.

Index Terms: Speech analysis, speaker variability, Astronaut

1. Introduction

Speech technology such as automatic speech recognition (ASR), keyword spotting, speaker identification (SID), speaker diarization and cognitive state detection has the potential of becoming an important part of future space missions [1–9]. Whereas simple command recognition system has been successfully tested on International Space Station [10], the applications of speech technology on more complicated task such as monitoring astronaut’s behavioral health and performance [11–14] requires a deep understanding of astronaut’s voice characteristic during space missions. For example, little research has considered how far astronaut speech production mechanisms change when they are experiencing physical stress such as weightlessness, high g-force, and an alternating pressure based environment which are common during space missions. Since training state-of-the-art ASR and SID systems for space applications depends highly on the data collected on the ground, the performance of those systems are expected to drop significantly if astronaut’s voice characteristics change in space.

The effect of physical stress on human voice has been widely investigated in previous studies [15–17]. These include the studies on speech under noise (Lombard effect) [15, 18], physical tasks stress [19], high g-force during air flight [20], or different air pressure such as in the deep sea environment [21]. However, to the best of our knowledge, no studies have yet attempted an analysis of human voice on subjects from Earth through an entire space mission.

The biggest challenge on the task of understanding human voice in space is the collection of audio data for analysis. After all, a small number of population has been traveled in space, while only 12 men in human history has been walked on the moon. In this study, we analyze the astronaut’s voice based on the historical recordings of air-to-ground communication of Apollo 11 as our first step towards this challenging problem. Specifically, the purpose of this study is to find out (1) whether astronaut’s voice characteristic changed when they were in space during Apollo 11, and (2) how did this affect the astronauts’ acoustic models which closely related to various speech application technologies.

In Sec. 2, we will give a brief description of the data used in our study. Sec. 3 will analysis the astronauts’ fundamental frequency as well as their acoustic models during Apollo 11.

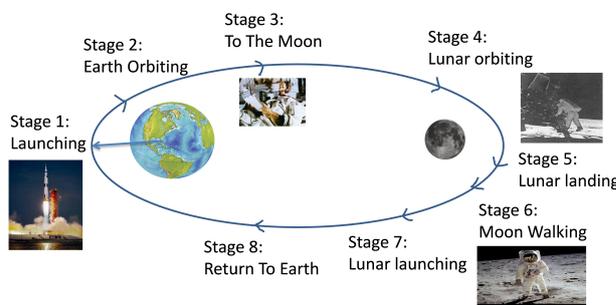


Figure 1: Overview of the timeline of Apollo 11 mission

2. Experiment design

By landing the first man on the moon, Apollo 11 (lasted 8 days 3 hours 18 minutes and 35 seconds) stands as one of mankind’s greatest achievements in the 20th century. A total of three astronauts participated in Apollo 11 missions. Among them, Neil Armstrong and Buzz Aldrin has walked on the Moon, while the third member, Michael Collins, did not land on the Moon. The speech data used in this study comes from air-to-ground communications of NASA Apollo 11 mission. The entire communications between astronauts and flight controller on the ground were recorded during the mission and being currently publicly available.

The Apollo 11 mission can be broadly separated into 8 stages (see Fig. 1) starting from the launching of spacecraft

	G-force	Environmental Pressure	Speech Samples			Speech Sources
			Armstrong	Aldrin	Collins	
Earth	1 g	14.7 PSI	7 min	6 min	6min	Postflight interview
Launch	1-4 g	5 PSI	1 min	N/A	N/A	Stage 1
Travel	zero	5 PSI	7 min	27 min	36 min	Stage 2,3,4
Lunar	1/6 g	5 PSI	N/A	2 min	4 min	Stage 6,7
Moon	1/6 g	3.5 PSI	5 min	5 min	N/A	Stage 8

Table 1: *Apollo 11 dataset used for analysis*

on Earth. In each of those stages, astronauts were experiencing various types and levels of stress. For example, astronauts were experiencing g-force with up to 4g during the launching of spacecraft, and 1/6g when they were on the moon (including landing and launching on the moon), and zero-g (weightless) other times during Apollo 11 mission. Another physical stress on astronauts were in terms of environmental pressure as the atmosphere during Apollo 11 was different from that is on Earth. For instance, the atmospheric pressure is approximately 14.7 PSI¹ on Earth at the sea level, 3.5 PSI pure oxygen while they were walking on the moon, and 5 PSI other times during Apollo 11 mission². Both g-force and environmental pressures could effect the respiratory system and therefore astronauts’ speech production mechanism.

To be able to classify various physical stress conditions, we categorize audios from all stages as well as those recorded on Earth into 5 conditions as described in Table. 1. The length of available speech samples for each condition is also listed. Note that all speech samples used in our experiments are pure speech signals with silence removed beforehand.

3. Analysis

3.1. Fundamental frequency

The fundamental frequency has been widely accepted as a reliable indication of altered voice characteristic. Previous studies have also reported that the fundamental frequency can be used for predicting speaker’s status under both physical and cognitive stress [15, 17, 19].

In this experiment, fundamental frequency was computed using WaveSurfer for every 10ms with an analysis window of 75ms. The range of fundamental frequency was set between 75 Hz to 350 Hz. The mean and standard deviation of fundamental frequency of all conditions across Apollo 11 mission are listed in Table 2. The lower panel of Fig. 2 are the distributions of fundamental frequency, while the upper panel shows the corresponding box plot.

From the results in Table. 2 we could observe that the mean and standard deviation of fundamental frequency of all astronauts are consistently higher during Apollo 11 mission (Launch, Space, Travel, and Moon) than on Earth. It is also noted that the fundamental frequency of Armstrong’s voice was significantly higher when he was on the moon compared to any other conditions. The relative increase of the Armstrong’s fundamental frequency is approximately of 35% when compared to his voice on Earth. Another interesting result not contained in Table. 2 is

¹PSI is an abbreviation for pound per square inch.

²Note that the atmospheric pressure at the top of Mount Everest is approximately 4.9 PSI.

	Armstrong		Aldrin		Collins	
	mean	std	mean	std	mean	std
Earth	114.3	18.17	102.5	16.1	105.7	17.2
Launch	137.4	36.3	N/A	N/A	N/A	N/A
Travel	130.4	25.2	114.0	22.0	124.5	23.4
Lunar	136.1	21.4	111.7	18.6	135.4	20.5
Moon	154.3	25.6	102.8	13.1	N/A	N/A

Table 2: *Mean and standard deviation of fundamental frequency across all conditions.*

that the fundamental frequency of Armstrong’s voice reached a peak of 160 Hz when he uttered the famous quote: ‘That’s one small step for [a] man, one giant leap for mankind’. While it is obvious that Armstrong’s fundamental frequency varied significantly when he was on the Moon, we do not observe the same effect on the other astronaut, Buzz Aldrin, who was also on the Moon.

In addition, Fig 2 shows that the distributions of fundamental frequency also varied within different phases of the Apollo 11 mission. The corresponding box plot shows the dispersion of fundamental frequency is much higher during Apollo 11 mission than on Earth.

3.2. Acoustical model

While fundamental frequency is important for excitation analysis, acoustic model analysis offers an opportunity to explore the speech production space with respect to phonemes. Acoustical model analysis is a parametric method of analyzing the variations of speaker’s vocal characteristic. It has two advantages compared to other vocal feature based analysis: (1) the result of analysis is closely related to many applications in speech technology such as ASR and SID, and (2) it does not require accurate transcription which is expensive both in terms of cost and time.

In this experiment, speech samples from each condition are broken into several 60-second chunks. For each 60-second speech chunk, 12 dimensional MFCC features are extracted and Gaussian mixture models were used to model their acoustical spaces. After modeling each speech chunk with a GMM, a cross comparison between models was achieved with a symmetric Kullback-Leibler (KL) divergence measure. The symmetric KL divergence is a popular and robust method for measuring the similarity between two probability density functions.

The results of cross comparison between 60-second chunks are plotted in Fig 3. As the symmetric KL divergence is used for computing the distance between two models, the resulting matrix of cross comparisons is also symmetric. It is observable

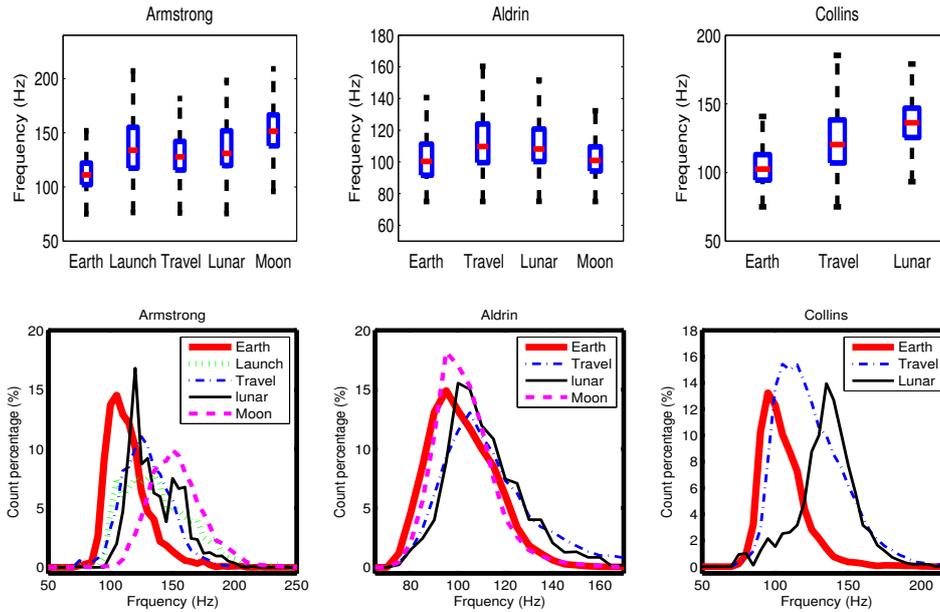


Figure 2: Fundamental frequency (F_0) across different physical stress conditions of astronauts during Apollo 11 mission. Upper panel is Box-and-Whisker plot indicating the median, quartiles, as well as smallest and greatest values of the distribution of F_0 . Lower panel is the distributions of F_0 without any prior assumption on underlying statistical distribution.

from Fig. 3 that the distance between models trained from Earth and other conditions during Apollo 11 mission (circled with yellow dashed line in Fig. 3) is obviously higher than the rest. This evidence indicates that the astronaut’s voice characteristic was significantly different during Apollo 11 mission compared to their speech when they were on Earth.

While it is obvious from Fig. 3 that astronauts acoustical models trained on Earth are highly different from those trained on conditions belong to Apollo 11 mission, it is not clear whether the astronauts’ acoustical models also varied within Apollo 11 by changing stress conditions. In order to investigate the effect of varying physical stress on astronauts’ acoustical models within Apollo 11 mission, the distances in each condition are averaged and then normalized using the average value of within condition distances. Therefore, a normalized distance between two conditions with a value larger than 1 indicates that the difference between corresponding two conditions is higher than the normal intra-speaker variabilities. Note that the distance related to the condition on Earth is not considered in this experiment. The result is listed in Table. 3.

From the result in Table. 3, we observe that the normalized distances of all pairwise Apollo conditions/phases have values higher than 1. This indicates that the difference of acoustical models between conditions are consistently higher than the average intra-speaker variabilities.

4. Conclusion

In this study, we performed an extensive study of the three astronaut voices recorded during Apollo 11 mission (i.e., 8 days, from July 20, 1969) and their voices on Earth. The results indicate that the mean, standard deviation, and dispersion of fundamental frequency increased consistently when astronauts were in space. In addition, acoustic models from astronaut voice during the Apollo 11 mission also varied compared to that from

	Launch	Travel	Lunar	Moon
Launch	0	1.36	1.2	1.46
Travel	1.36	1.142	1.06	1.29
Lunar	1.2	1.06	0	1.05
Moon	1.46	1.29	1.05	0.54

Table 3: Normalized distance between speaker GMM models from different conditions/phases during Apollo 11.

Earth. We attribute those difference to an increase in physical and cognitive stress as well as alternate oxygen/air pressure within the space environments.. Moreover, it is observed in this study that variation in voice characteristic also occurs within different stages of Apollo 11 mission.

The above result is important for further applications of speech technology for extended space missions. It suggests that the corresponding acoustical model widely employed in various speech technologies must be capable of adapting to varying conditions if humans are considering extended periods of human exploration in space.

5. Acknowledgments

This research was supported by National Science Foundation (NSF) under Grant 1219130.

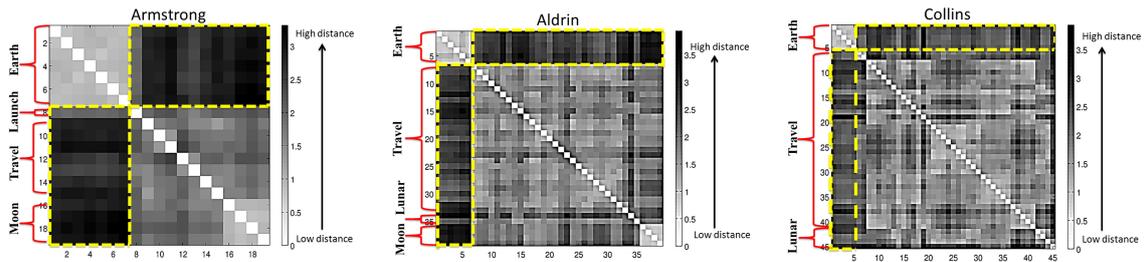


Figure 3: Cross comparison of acoustical models trained from different 60-second chunks using KL divergence.

6. References

- [1] T. Hasan, S. O. Sadjadi, G. Liu, N. Shokouhi, H. Boril, and J. H. L. Hansen, "CRSS systems for 2012 NIST speaker recognition evaluation," in *Proc. of ICASSP*, 2013, pp. 6784–6786.
- [2] J. H. L. Hansen, S. E. Bou-Ghazale, R. Sarikaya, and B. Pellom, "Getting started with SUSAS: a speech under simulated and actual stress database.," in *EUROSPEECH*, 1997, vol. 97, pp. 1743–46.
- [3] J. H. L. Hansen, R. Huang, B. Zhou, M. Seadle, Deller J. R., A. R. Gurijala, M. Kurimo, and P. Angkititrakul, "Speechfind: Advances in spoken document retrieval for a national gallery of the spoken word," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 5, pp. 712–730, 2005.
- [4] A. Sangwan and J. H. L. Hansen, "Keyword recognition with phone confusion networks and phonological features based keyword threshold detection," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, 2010, pp. 711–715.
- [5] G. Liu, C. Yu, A. Misra, N. Shokouhi, and J. H. L. Hansen, "Investigating State-of-the-Art Speaker Verification in the Case of Unlabeled Development Data," *Proc. Odyssey speaker and language recognition workshop, Joensuu, Finland*, 2014.
- [6] C. Yu, G. Liu, S. Hahm, and J. H. L. Hansen, "Uncertainty Propagation in Front End Factor Analysis For Noise Robust Speaker Recognition," *Proc. ICASSP, Florence, Italy*, 2014.
- [7] C. Yu, G. Liu, and J. H. L. Hansen, "Acoustic Feature Transformation using UBM-based LDA for Speaker Recognition," *Proc. Interspeech, Singapore*, 2014.
- [8] C. Yu, K. K. Wójcicki, M. T. Johnson, P. C. Loizou, and J. H. L. Hansen, "A new mask-based objective measure for predicting the intelligibility of binary masked speech," *Proc. ICASSP, Vancouver, Canada*, 2013.
- [9] C. Yu, K. K. Wójcicki, M. T. Johnson, P. C. Loizou, and J. H. L. Hansen, "Evaluation of the importance of time-frequency contributions to speech intelligibility in noise," *The Journal of the Acoustical Society of America*, vol. 135, no. 5, pp. 3007–3016, 2014.
- [10] M. Rayner, B. A. Hockey, J. Renders, N. Chatzichrisafis, and K. Farrell, "Spoken dialogue application in space: The Clarissa procedure browser," in *Speech Technology*, pp. 221–250. Springer, 2010.
- [11] E. Salas, J. E. Driskell, and T. Driskell, "Using real-time lexical indicators to detect performance decrements in spaceflight teams: A methodology to dynamically monitor cognitive, emotional, and social mechanisms that influence performance," in *NASA Human Research Program Investigators' Workshop*, 2014, p. 3089.
- [12] S. K. Baard, S. Lorenz, M. Pearce, D. B. C. Kermond, S. Golden, S. Biswas, C.-H. Chang, and S. W. J. Kozlowski, "Monitoring team dynamics through physical and physiological indicators," in *NASA Human Research Program Investigators' Workshop*, 2014, p. 3176.
- [13] A. Sangwan, L. Kaushik, C. Yu, J. H. L. Hansen, and D. W. Oard, "Houston, We have a solution: Using NASA Apollo Program to advance Speech and Language Processing Technology," in *Proc. of INTERSPEECH*, 2013, pp. 1135–1139.
- [14] D. Oard, A. Sangwan, and J. H. L. Hansen, "Reconstruction of apollo mission control center activity," in *the Proceedings of the First Workshop on the Exploration, Navigation and Retrieval of Information in Cultural Heritage, ENRICH*, 2013.
- [15] J. H. L. Hansen, "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," *Speech communication*, vol. 20, no. 1, pp. 151–173, 1996.
- [16] S. A. Patil and J. H. L. Hansen, "Speech under stress: Analysis, modeling and recognition," Springer-Verlag, 2007.
- [17] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 3, pp. 201–216, 2001.
- [18] J. H. L. Hansen, "Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 4, pp. 598–614, 1994.
- [19] K. W. Godin and J. H. L. Hansen, "Analysis of the effects of physical task stress on the speech signal," *The Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3992–3998, 2011.
- [20] C. A. Simpson, "Speech variability effects on recognition accuracy associated with concurrent task performance by pilots," *Ergonomics*, vol. 29, no. 11, pp. 1343–1357, 1986.
- [21] G. Fant and B. Sonesson, "Speech at high ambient air-pressure," *Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm*, 1964.