Interspeech 2018
2-6 September 2018, Hyderabad

The Effect of Exposure to High Altitude and Heat on Speech Articulatory Coordination

James R. Williamson[1], *Thomas F. Quatieri*[1], *Adam Lammert*[1],
Katherine Mitchell[2], *Katherine Finkelstein*[2], *Nicole Ekon*[2], *Caitlin Dillon*[2],
Robert Kenefick[2], *Kristin Heaton*[2]

[1]MIT Lincoln Laboratory
[2]USARIEM

jrw@ll.mit.edu

Abstract

The effects of altitude and heat on speech articulatory coordination following exercise and approximately three hours of exposure are explored. Recordings of read speech and free response speech before and after exercise in moderate altitude, moderate heat, and both moderate altitude and heat are analyzed using features that characterize articulatory coordination. It is found that 1) moderate altitude causes small changes and moderate heat negligible changes to articulatory coordination features after brief exposure prior to exercise; 2) moderate altitude and heat produce similar large feature changes following exercise and longer exposure; 3) moderate altitude and heat produce larger feature changes in combination than individually immediately following exercise. Finally, using cross-validation training of a statistical classifiers, the features are sufficient to classify the four experimental conditions with an accuracy of 0.50, and to detect the presence of any one of the experimental conditions with an accuracy of 0.90.

Index Terms: speech analysis, articulatory coordination, altitude and heat effects

1. Introduction

In many regions of the world there are environments in which both hot ($> 30^o$ C) and moderate altitude ($> 3,000$ m) ambient conditions co-exist, such as South East Asia. However, to our knowledge there has never been a scientific study of the combined impact of these environmental factors on physiological and cognitive function prior to, during, and after exercise. In this paper the goal is to assess the effect of aerobic exercise under acute heat and altitude conditions on fine motor control involved in speech articulatory coordination,

Military operational environments present service members with numerous mental and physical challenges, which can include extreme heat, altitude, or both in combination. These conditions can degrade cognitive function due to both reversible (e.g., cognitive workload) and irreversible (e.g., neural injury/cell loss) causes. Although studies examining the effects of moderate altitude exposure on cognitive function have yielded inconsistent findings [1, 2], research evaluating performance at high altitude (e.g., $> 4,500$ m) have noted degraded function across many key cognitive domains, particularly motor, memory, language and executive function in adults (e.g.,[3, 4]).

Speech markers that may have particular relevance for detection of cognitive load, fatigue, or performance include prosody and pitch variation [5], grammar [6], and formant structure [7]. In particular, measures of articulator coordination [8] based on formant structure and dynamics have shown excellent sensitivity for detecting depression [8, 9], cognitive decline in mild TBI [10], dementia in the elderly [11], cognitive decline following heavy cognitive workload [12, 13], and severity of Parkinson's disease [14]. Given the effects of altitude on motor function, these speech-based measures are particularly appropriate tools for assessment of possible changes in cognition due to altitude exposure.

Changes in speech have been studied specifically with respect to hypoxia. Lieberman [15] and Cymerman [16], for example, have shown that voice onset time (VOT), the duration between a consonant burst and onset of the following vowel, as well as the duration of the following vowel, lengthen under high altitude conditions. These studies used a speech database consisting of monosyllabic voiced/voiceless word pairs such as "to" vs "do" and "cod" vs "god" in which the voice onset time is a distinguishing characteristic. Lieberman and Cymerman additionally considered cognitive deficits in conjunction with speech changes. Speech changes were correlated with increased difficulty performing set-shifting cognitive tasks (such as "odd man out" card sorting) and increased time required for simple sentence comprehension. Experiments were performed under both real (Mt. Everest) and simulated (hypobaric chambers) conditions. More recent work by Kiss [17] suggests that a general loss of speech clarity can be expected in individuals exposed to relatively long periods of reduced oxygen levels.

In Section 2, a new data collection is described that probes the interactions of moderate altitude and moderate heat, with and without physical exertion, on speech articulatory coordination. Speech recordings were obtained under hot or moderate altitude, or under combined hot and moderate altitude conditions, both before and after a sustained aerobic exercise task (45 min duration on a cycle ergometer), within the first 2 hours of exposure in a hypobaric chamber with temperature and humidity controls. In Section 3, articulatory coordination features are described that use time varying resonances of the vocal tract as indicators of coordination among speech articulators. In Section 4, changes in the articulatory coordination features due to different stimulus combinations are quantified. In addition, the net discriminative value of the articulatory coordination features is assessed based on the ability to classify experimental conditions in cross-validation. In Section 5, the relevance of these findings, and our proposed path forward, is discussed.

2. Data Collection

The data collection was conducted to enable investigation of effects of altitude and heat exposure on physical and cognitive performance, including fine motor control in speech production. The subject population consists of thirteen healthy and fit young

*This material is based upon work supported by the Department of the Army under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Army.

10.21437/Interspeech.2018-2372

Table 1: *Subject demographics*

Description	Mean	SD
Age (yrs)	25.3	6.1
Baseline Wt (kg)	81.2	16.8
Height (m)	1.8	0.1
% Body fat	22.2	5.5
SL VO_2 Peak (ml/kg/min)	41.9	4.4

Table 2: *Experimental conditions*

Condition	Altitude	Temp.	Rel. Humid.
1	250 m	20 deg C	30-50%
2	250 m	35 deg C	30%
3	3,000 m	20 deg C	30-50%
4	3,000 m	35 deg C	30%

men and women. Table 1 provides demographic characteristics for the study sample. For each experimental condition, the experimental protocol involved a dwell time of three hours within a climate controlled hyperbaric chamber, with 45-minutes of aerobic exercise following the first hour. Four different experimental conditions were conducted in which altitude and heat were varied in the chamber. In each condition, a standardized speech protocol was administered at five time points: one before entering the chamber, three inside the chamber (one before and two following aerobic exercise), and one after leaving the chamber.

The four experimental conditions are listed in Table 2. Condition 1 is the control condition and Conditions 2-4 are the experimental conditions that probe the effects of moderate altitude, moderate heat, and combined altitude and heat.

Figure 1 diagrams the experimental protocol used for each condition. A set of speech recordings take place less than one hour after entering the chamber. At the one hour mark the subject does 45 minutes of aerobic exercise on a cycle ergometer. The experiment is designed to probe the effect on fine speech motor control of exposure to high altitude and/or high heat conditions following physical exertion and prolonged exposure.

The first and third set of speech recordings in the chamber comprise the full speech protocol, which includes a read passage (rainbow), two image response free speech passages, repetitions of pataka, and several held vowels. The second speech recording in the chamber includes only the read passage and a single image free response passage.

The analysis technique used in this paper is designed to characterize articulatory dynamics in read or free speech with a duration of several seconds or more. Therefore the analysis in this paper is restricted to the read passage and the two free speech passages. This results in eight speech recordings inside the chamber in each experimental condition.

3. Articulatory Coordination Features

Articulatory coordination features are designed to capture levels of dynamical interrelations among the multiple articulators involved in speech production. More specifically, the features are derived from the eigenspectra of correlation matrices, which are constructed from time series of the first three formant frequencies, with expanded dimensionality based on time delay embedding. The resulting eigenvalues capture the shape of the

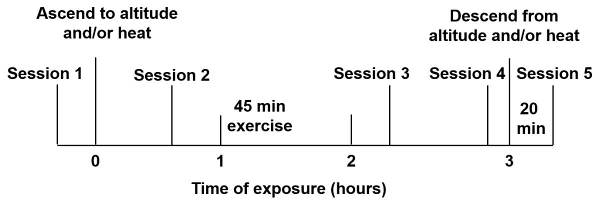

Figure 1: *Time course of each experimental condition. Sessions 1-5 are the speech recording sessions.*

high-dimensional scatter distribution of the multiple formant channels at multiple relative time lags. Previously, it was found that scatter distributions with a more spherical quality (i.e., having larger values in the lower rank eigenvalues) reflect healthier states, such as lower levels of depression, Parkinson's Disease, and less impaired cognition [8, 9, 10, 11, 14]. This multivariate feature approach was first introduced for analysis of spatiotemporal coordination among EEG signals for predicting epileptic seizures [18, 19].

3.1. Audio formant frequencies

Vocal tract resonances ("formants") contain information about speech dynamics based on the rates and relative timings of speech articulator movements, which shape the vocal tract. A formant tracking algorithm based on Kalman filtering is used to obtain smooth estimates of the first three resonant frequencies over time [20]. Formant frequencies are extracted every 10 ms from the audio signal. Embedded in the formant tracking algorithm is a voice-activity detector that allows a Kalman smoother to smoothly coast through non-speech regions. Post processing is also applied, in which estimates of the third formant frequency that are above a threshold of 4.5k Hz are truncated.

3.2. Articulatory coordination features

High-level articulatory coordination features are extracted from the low-level formant time series data. In this feature approach, channel-delay correlation matrices are computed from low-level multi-channel signals using time-delay embedding at multiple different delay scales. The features consist of the eigenspectra from the correlation matrices.

The time delay embedding is done at four different delay scales, with delay spacings of 1, 3, 7, and 15 frames, which correspond to 10, 30, 70, and 150 ms time delays. There are 15 time delays per scale, ranging from zero delay up to 14 times the scale-depending delay. Thus, in each scale, time-embedded correlation matrix is created with dimensionality (45 x 45), due to three formant channels and 15 time delays per channel. From each correlation matrix a 45-dimensional rank ordered eigenspectrum is computed, which characterizes the within-channel and cross-channel distributional properties of the formant time series.

An important consideration in using audio features at moderate or high altitudes is the impact of physical atmospheric conditions on the production of speech. At high altitudes, the speed of sound is reduced compared to sea level, which directly affects the resonant properties of the vocal tract. The U.S. Standard Atmosphere report [21] predicts a drop in the speed of sound of approximately 5% at altitudes of 15,000 ft. which should be accompanied by a corresponding rise in vocal tract resonances of 5%. Because the articulatory coordination features used in this study are derived using correlation coefficients, in which absolute formant magnitudes are factored out, the features are invariant to shifts in absolute frequency values due to altitude.

4. Results

4.1. Feature changes

At a given delay scale, the changes of the articulatory features, which are correlation matrix eigenvalues, are assessed at each eigenvalue rank. The eigenvalues are first normalized (z-scored) so that the changes are expressed in standard units. This normalization is done based on feature statistics from Session 1 of all subjects, under all experimental conditions, for a given type of recording (e.g., read speech or free speech). Session 1 is used because it precedes exposure to the heat and altitude manipulations.

In each experimental condition (Conditions 2-4), changes in the eigenvalue features are calculated relative to the eigenvalues of the control condition (Condition 1), which serves as a baseline. This is done for read speech (Rainbow passage) and the first free speech response to a visual image (”Describe what you see in this photo”), which are available in all three sessions. Feature differences for Session 5 (return to baseline conditions) were also calculated. These differences were negligible, and are not shown here.

Figure 2 shows the Cohen’s D effect sizes of the eigenvalue feature changes in the first delay scale for Sessions 2 (top), 3 (middle) and 4 (bottom), on the two speech recordings. For each session the effect sizes are computed using 18 data points (nine subjects and two recordings per subject). Four of the 13 subjects are not used in this analysis due to missing data in one or more of the three analyzed sessions. With $n = 18$, p values of 0.05 and 0.01 correspond to effect sizes of 0.52 and 0.69, respectively.

Notice that small eigenvalues, particularly those between indexes 7 and 26, have positive changes, on average, in all three experimental conditions. These effects are much larger in session 4, following exercise and prolonged exposure in heat and/or altitude, than they are in session 2, which occurs prior to exercise.

Figure 2 also reveals a different time course of feature changes for heat alone (Condition 2) than for altitude alone (Condition 3). Heat produces a negligible initial effect. But, with exercise and time the heat effect becomes similar in size to the altitude effect. Thus, the effect of heat on speech dynamics begins more slowly but accumulates more due to exercise and prolonged exposure than the effect of altitude. The combined effect of altitude and heat (Condition 4) also shows this strong cumulative effect.

In Figure 3 these positive effects are summarized by computing effect size from the average of 20 neighboring normalized eigenvalues in the first delay scale. The rank indices 7-26 are chosen because these are the (20-neighborhood) indices that yield the largest average effect size in the three experimental conditions in the three sessions. The trends revealed in Figure 3 are insensitive to the precise size of the neighborhood of eigenvalues that are averaged. Notice that combined altitude and heat exposure produce a stronger effect immediately following exercise than either exposure individually.

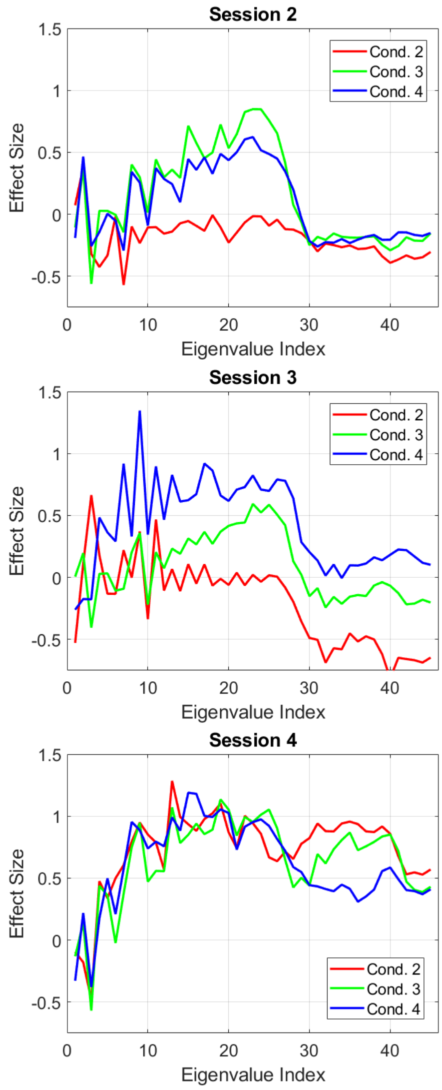

Figure 2: *Effect sizes for normalized eigenvalues from the first formant delay scale in the three experimental conditions relative to the control condition, in Session 2 (top), Sesssion 3 (middle), and Session 4 (bottom).*

4.2. Classifying experimental condition

The features in the three experimental sessions, derived from seven speech recordings, were used to train a system to classify the experimental condition that gave rise to the speech recordings. This was done to quantify and summarize the total generalizeable discriminative information that is available from the feature vectors at all four delay scales and from all experimental speech recordings.

The classification method is as follows. For each speech recording and each formant delay scale, a statistical model was fit to the formant-based feature distribution. First, The feature elements were z-scored based on all of the recordings in the training set. Next, PCA was used to reduce the dimensionality from 45 dimensions (the number of eigenvalues in each correlation matrix) to five dimensions (a typical number of principal component features used in previous studies). Then, each PCA feature was z-scored based on the feature means and standard

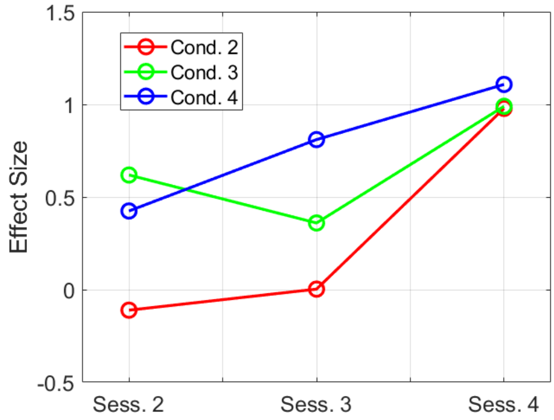

Figure 3: *Effect sizes of normalized eigenvalues, averaged across indices 7-26 in the first formant delay scale.*

Table 3: *Four-class confusion matrix showing probability of predicting each experimental condition as a function of the true condition. The overall classification accuracy is 0.50.*

True Cond.	Pred. Cond. 1	2	3	4
1	0.77	0.08	0.08	0.08
2	0.08	0.15	0.31	0.46
3	0.08	0.23	0.31	0.38
4	0.0	0.15	0.08	0.77

deviations on the training set. Finally, a multivariate normal distribution was fit to this 5-dimensional normalized PCA feature space. For regularization, a constant of one was added to the diagonal of the covariance matrix.

Because there are eight within-chamber speech recordings per experimental condition, and four formant delay scales, a total of 32 Gaussian models were constructed for each experimental condition class model. Leave-one-subject-out cross-validation for all 13 subjects was used to classify which experimental condition gave rise to the seven within-chamber speech recordings of a test subject. In each cross-validation fold, the normalizations and PCA transforms described above were applied to the test features. Then, the likelihoods on the test data were computed for the 32 class models for each of the four experimental condition classes. The 32 log-likelihoods for each class were summed together, and the class with the highest net log-likelihood was selected as the predicted experimental condition.

Table 3 shows the resulting four-class confusion matrix computed across all 13 subjects. The average accuracy is 50%. The control condition (Condition 1) is easily distinguished from the three experimental conditions. However, correct classification among the three experimental conditions is difficult.

In Tables 4-6, marginal classifications are computed by appropriately combining the class predictions from the full confusion matrix. Table 4 shows that the presence of an experimental condition (heat, altitude, or both) is detected with very high accuracy (90%). Tables 5 and 6 show that high altitude and high temperature are also detected with high accuracy (65 and 69%, respectively).

Table 4: *Two-class confusion matrix for detecting any of the three experimental conditions versus the control condition. The overall classification accuracy is 0.90.*

True Cond.	Pred. Cond. 1	2, 3 or-4
1	0.77	0.23
2, 3 or 4	0.05	0.95

Table 5: *Two-class confusion matrix for detecting high altitude. The overall classification accuracy is 0.65.*

True Cond.	Pred. Cond. 1 or 2	3 or 4
1 or 2	0.54	0.46
3 or 4	0.23	0.77

Table 6: *Two-class confusion matrix for detecting high heat. The overall classification accuracy is 0.69.*

True Cond.	Pred. Cond. 1 or 3	2 or 4
1 or 3	0.62	0.38
2 or 4	0.23	0.77

5. Discussion

This paper presents a new study that explores the effects of high altitude and heat on speech with and without physical exertion. Speech articulatory coordination is characterized from recordings of read speech and free speech. It is found that altitude and altitude combined with heat has a moderate effect prior to exercise. Following exercise and with longer exposure duration, all three exposure conditions (altitude, heat, and altitude and heat combined) have a large and similar type of effect on articulatory coordination. Combined exposure to altitude and heat produces the largest effect.

The effect of altitude and heat on the articulatory coordination features is counterintuitive, Based on previous findings [8, 9, 10, 11, 14], we would expect the physical stress of these conditions to reduce articulatory coordination, manifested in a reduction in small eigenvalues. However, we see the oppo-site. There are two conclusions one can draw from these find-ings: 1) the experimental manipulations temporarily improve neuromotor processing, which results in greater speech artic-ulatory coordination; 2) larger values in low rank eigenvalues is not a specific indicator of effective articulatory coordination, but rather an indicator of greater complexity in articulatory dynamics. This increased complexity may indicate more efficient neuromotor processing that produces more intelligible and dynamic speech. Alternatively, it may indicate greater variability in articulatory dynamics, due to degraded neuromoter processing, which results in less intelligible speech. A future research direction is to distinguish between these two interpretations based on a more detailed analysis of the current dataset. Other future areas include analysis of ongoing data collections with heat and altitude manipulations, in which speech recordings are accompanied by a battery of cognitive tests, expanding our vocal feature suite to prosodic-, source-, and phonetic-based features, and adding features derived from our companion facial video recordings.

6. References

[1] B. Shukitt-Hale, L. E. Banderet, and H. R. Lieberman, “Relationships between symptoms, moods, performance, and acute mountain sickness at 4,700 meters.” *Aviation, space, and environmental medicine*, 1991.

[2] G. E. Adam, C. S. Fulco, and S. R. Muza, “Multi-task performance at sea-level and high altitude,” ARMY RESEARCH INST OF ENVIRONMENTAL MEDICINE NATICK MA, Tech. Rep., 2008.

[3] J. Virués-Ortega, G. Buela-Casal, E. Garrido, and B. Alcázar, “Neuropsychological functioning associated with high-altitude exposure,” *Neuropsychology review*, vol. 14, no. 4, pp. 197–224, 2004.

[4] J. Virués-Ortega, E. Garrido, C. Javierre, and K. C. Kloezeman, “Human behaviour and development under high-altitude conditions,” *Developmental Science*, vol. 9, no. 4, pp. 400–410, 2006.

[5] R. M. Shiffrin and W. Schneider, “Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory.” *Psychological review*, vol. 84, no. 2, p. 127, 1977.

[6] M. A. Khawaja, F. Chen, and N. Marcus, “Analysis of collaborative communication for linguistic cues of cognitive load,” *Human factors*, vol. 54, no. 4, pp. 518–529, 2012.

[7] S. E. Lively, D. B. Pisoni, W. Van Summers, and R. H. Bernacki, “Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences,” *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2962–2973, 1993.

[8] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, “Vocal biomarkers of depression based on motor incoordination,” in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*. ACM, 2013, pp. 41–48.

[9] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, “Vocal and facial biomarkers of depression based on motor incoordination and timing,” in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014, pp. 65–72.

[10] B. S. Helfer, T. F. Quatieri, J. R. Williamson, L. Keyes, B. Evans, W. N. Greene, T. Vian, J. Lacirignola, T. Shenk, T. Talavage *et al.*, “Articulatory dynamics and coordination in classifying cognitive change with preclinical mtbi,” in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[11] B. Yu, T. F. Quatieri, J. R. Williamson, and J. C. Mundt, “Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers,” in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[12] T. F. Quatieri, J. R. Williamson, C. J. Smalt, T. Patel, J. Perricone, D. D. Mehta, B. S. Helfer, G. Ciccarelli, D. Ricke, N. Malyska *et al.*, “Vocal biomarkers to discriminate cognitive load in a working memory task,” in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[13] T. F. Quatieri, J. R. Williamson, C. J. Smalt, J. Perricone, T. Patel, L. Brattain, B. Helfer, D. Mehta, J. Palmer, K. Heaton *et al.*, “Multimodal biomarkers to discriminate cognitive state,” *The Role of Technology in Clinical Neuropsychology*, p. 409, 2017.

[14] J. R. Williamson, T. F. Quatieri, B. S. Helfer, J. Perricone, S. S. Ghosh, G. Ciccarelli, and D. D. Mehta, “Segment-dependent dynamics in predicting parkinson’s disease,” in *Sixteenth annual conference of the international speech communication association*, 2015.

[15] P. Lieberman, A. Morey, J. Hochstadt, M. Larson, and S. Mather, “Mount everest: a space analogue for speech monitoring of cognitive deficits and stress,” *Aviation, space, and environmental medicine*, vol. 76, no. 6, pp. B198–B207, 2005.

[16] A. Cymerman, P. Lieberman, J. Hochstadt, P. B. Rock, G. E. Butterfield, and L. G. Moore, “Speech motor control and acute mountain sickness.” *Aviation, space, and environmental medicine*, vol. 73, no. 8, pp. 766–772, 2002.

[17] G. Kiss, D. Sztahó, K. Vicsi, and A. Golemis, “Connection between body condition and speech parameters-especially in the case of hypoxia,” in *Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on*. IEEE, 2014, pp. 333–336.

[18] J. R. Williamson, D. W. Bliss, and D. W. Browne, “Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial eeg,” in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 665–668.

[19] J. R. Williamson, D. W. Bliss, D. W. Browne, and J. T. Narayanan, “Seizure prediction using eeg spatiotemporal correlation structure,” *Epilepsy & Behavior*, vol. 25, no. 2, pp. 230–238, 2012.

[20] D. D. Mehta, D. Rudoy, and P. J. Wolfe, “Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking,” *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1732–1746, 2012.

[21] N. Sissenwine, M. Dubin, and H. Wexler, “The us standard atmosphere, 1962,” *Journal of Geophysical Research*, vol. 67, no. 9, pp. 3627–3630, 1962.