Articulation entropy: An unsupervised measure of articulatory precision

Yishan Jiao, Visar Berisha, Julie Liss, Sih-Chiao Hsu, Erika Levy, and Megan McAuliffe

Abstract

Articulatory precision is a critical factor that influences speaker intelligibility. In this paper, we propose a new measure we call 'articulation entropy' that serves as a proxy for the number of distinct phonemes a person produces when he or she speaks. The method is based on the observation that the ability of a speaker to achieve an articulatory target, and hence clearly produce distinct phonemes, is related to the variation of the distribution of speech features that capture articulation - the larger the variation, the larger the number of distinct phonemes produced. In contrast to previous work, the proposed method is completely unsupervised, does not require phonetic segmentation or formant estimation, and can be estimated directly from continuous speech. We evaluate the performance of this measure with several experiments on two data sets: a database of English speakers with various neurological disorders and a database of Mandarin speakers with Parkinson's disease. The results reveal that our measure correlates with subjective evaluation of articulatory precision and reveals differences between healthy individuals and individuals with neurological impairment.

Index Terms

articulatory precision, phonemic inventory, entropy, unsupervised estimation, pathological speech

Y. Jiao, V. Berisha, and J. Liss are with the Department of Speech and Hearing Science, Arizona State University, United States. V. Berisha has a joint appointment in the School of Electrical, Computer, and Energy Engineering, Arizona State University, United States. S. C. Hsu and E. Levy are with the Department of Biobehavioral Sciences, Teachers College, Columbia University, United States. M. McAuliffe is with the Department of Communication Disorders and New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, New Zealand. This work was partially supported by a grant from NIH (1R21DC013812).

I. Introduction

Articulatory precision, or the accuracy with which articulators achieve their targets, varies naturally in everyday speaking, depending on the situation (casual versus formal communication settings), and the physical and psychological state of the speaker (e.g., fatigued versus excited) [1], [2]. However, progressive and unremitting reductions in articulatory precision can be a sign of underlying neurological disease [3], [4], [5]. For this reason, it is useful to quantitatively characterize articulatory precision as a potential indicator for neurological health. In this paper we define the precision of articulation as the size of a speaker's "working phonemic inventory" - the number of distinct phonemes a person produces when he or she speaks. To that end, we propose a method that estimates this value from continuous speech samples.

Estimating articulatory precision from speech acoustics traditionally has focused on specific aspects of vowel and consonant production. The vowel space area (VSA) is a commonly used metric in the evaluation of pathological speech. It is defined as the area of the quadrilateral spanned by the first and second formants of the 4 corner vowels /a, æ, i, u/ [6], [7]. This metric can be interpreted as a measure of articulatory excursions and separability between distinct vowel targets. It has been shown that dysarthric speakers have relatively compressed VSA when compared against healthy individuals [8], [9], [10], [11]. The VSA is typically estimated from isolated words that contain the corner vowels (e.g. hit, hat, hut, hot) with manual labeling of individual phonemes, which can be cumbersome, time consuming, and error prone; a notable exception is the paper by Sandoval et al. that measures the VSA directly from continuous speech based on formant estimation of voiced speech segments (including vowels and voiced consonants) [12]. Related to the VSA, Sapir and colleagues have proposed two other measures of articulatory precision called the Vowel Articulation Index (VAI) and, its inverse, the Formant Centralization Ratio (FCR) [13]. These methods are less sensitive to inter-speaker variability and more sensitive to vowel formant centralization [14]. While these measures have shown to be useful in a number of applications, they require precise formant estimation and hand labeling of individual vowels. Furthermore, vowels only account for a part of a speech signal. It is known that intelligibility also strongly depends on consonant production [15], [16]. Acoustic measures of consonant precision traditionally have targeted spectral characteristics of stops, fricatives and affricates [17], [9].

In this paper, we propose an unsupervised metric that considers both vowel and consonant production and does not require hand labeling or formant estimation. Furthermore, it is a language

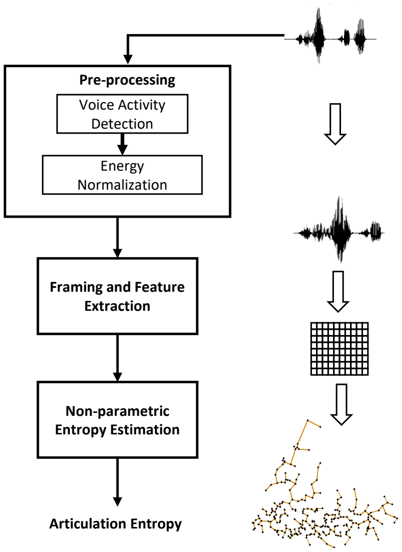

Fig. 1. A framework for calculating articulation entropy.

independent measure and can be applied to continuous speech samples. In information theory, *entropy* measures the expected amount of information contained in a message [18]. We extend this idea to the acoustic representation of speech production. Here we consider the distribution of sounds produced by an individual and use a non-parametric estimate of the entropy of this distribution to characterize the speaker's working phonemic inventory - we call this measure the 'articulation entropy' of the speaker. We evaluate the algorithm on speech samples from dysarthric English and Chinese patients and show that articulation entropy correlates with perceptual impressions of articulatory precision.

II. Articulation entropy

Consider a speech signal, $x_i(t)$, where i is the index of the speaker, with sampling rate F_s. We assume that this signal can be partitioned into N_i equal length segments (between 40*ms* to 200*ms*). For each frame, a D-dimensional feature vector is extracted and denoted by the random variable $\mathbf{z}_i \in \mathbf{R}^{D\times 1}$, sampled from an unknown continuous distribution $f^i_{\mathbf{z}}(\mathbf{z})$. We posit that the entropy of $f^i_{\mathbf{z}}(\mathbf{z})$ is a proxy for the working phonemic inventory as it captures the diversity of sounds produced as a person speaks. The framework to calculate articulation entropy is shown in Figure 1. It consists of a pre-processing step that normalizes the intensity of the speech signal, followed by feature extraction, and a non-parametric entropy estimation step. We describe these steps below.

A. Pre-processing

Speech samples are first passed through a voice activity detection (VAD) algorithm to exclude all silences and pauses during speech. We use the VAD algorithm described in [19]. To remove energy-dependence effects on the variation in the speech features, we normalize the intensity of the speech signal after VAD. To normalize, we select a speech sample as a reference and normalize all of the other speech samples from all speakers to the energy level of the reference sample.

B. Feature extraction

Two window lengths are adopted to capture the short-term speech acoustics and the long-term articulatory movement: we use a short term 20*ms* frame and a longer analysis window that ranges between 40*ms* to 200*ms*. The speech signal is initially split into 20*ms* frames (with 10*ms* overlap) and a feature vector is extracted for each frame. The 20*ms* feature vectors are concatenated into a longer super-vector that captures the articulatory motion. For example, for a longer analysis window of 100*ms*, features from 10 consecutive (overlapping) 20*ms* frames are concatenated into a single vector. Previous studies reveal that average phoneme duration can vary between 40*ms* to 200*ms* [20]; therefore, in the results section we examine different analysis window lengths that span this range.

We evaluate our algorithm with four families of spectral features: (1) the log-magnitude of the FFT of each frame (2) the Mel-frequency cepstral coefficients (MFCCs), (3) the linear prediction coding (LPC) envelope, and (4) the mel-spectrum with cubic root compression features (MelRoot3) [21]. The performance of the algorithm using these features is reported in Section III-A1.

C. Entropy estimation

Given a speech sample from speaker i, the features are extracted as described in the previous section, then stacked in a single multi-dimensional feature matrix $\mathbf{Z}^i \in \mathbb{R}^{N_s^i \times D}$. where N_s^i is the number of long segments extracted from speaker i's speech sample. This feature matrix can be interpreted as a set of N_s^i samples of the unknown D-dimensional continuous distribution $f_{\mathbf{z}}^i(\mathbf{z})$. We associate the larger variation in the features with a larger working phonemic inventory. We use the entropy of the distribution from which these features are sampled to measure this variation.

Typically, estimating entropy requires complete knowledge of the data distribution. For example, a Gaussian distribution is often assumed and the closed-form entropy estimator is used. However, it is known that speech does not follow a Gaussian distribution [22], [23]. In fact, for our application, the underlying distribution is completely unknown. To estimate information-theoretic parameters in these scenarios a non-parametric approach can be used [24], [25], [26]. We estimate the Rényi entropy [27] directly from the data samples without having to make assumptions about the underlying data distribution. The Rényi entropy of the distribution is defined as

$$H_{\alpha}(f) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{N} f^{\alpha}(\mathbf{z}_i), \tag{1}$$

where α is a user-selected parameter between 0 and 1 [1]. The authors in [26], showed that the Rényi entropy can be estimated directly from a set of samples using a graph theoretic approach. From the samples, a fully connected Euclidean graph can be calculated. From the graph, the minimal spanning tree (MST) can be derived. The authors in [26] showed that the sum of the edge weights of the MST can be used to estimate the entropy. For more details on the estimator please refer to [26].

We illustrate this procedure with an example. The hypothesis is that when two speakers read the same content, the distribution of acoustic features for the speaker who has more precise articulation should have larger variation, and a larger entropy, than that of the speaker who has imprecise articulation. Consider features extracted from speech samples produced by two individuals who read the same content: One has mild dysarthria, and the other has severe dysarthria. For each frame, we extract 13-dimensional MelRoot3 features from 20*ms* frames, then concatenate these to form a 100*ms* analysis window (per section II-B). We reduce the dimension of this data using principal components analysis (PCA) to 2 for visualization [28]. In the actual implementation of the algorithm, much higher-dimensional feature sets are used. In Figure 2 we show two graphs, with the nodes representing a 100*ms* 2-D feature and the edges representing the Euclidean edges of the MST. In each figure, we also show the estimate of the entropy obtained per the method in [26]. From the figure, it is clear that the speaker with mild dysarthria (right) has a much higher articulation entropy than the speaker with severe dysarthria (left). Readers are invited to listen to the supplementary speech recordings as evidence of the perceptual differences in articulation.

[1] We set $\alpha = 0.99$ in the experiment, which approximates the commonly used Shannon entropy.

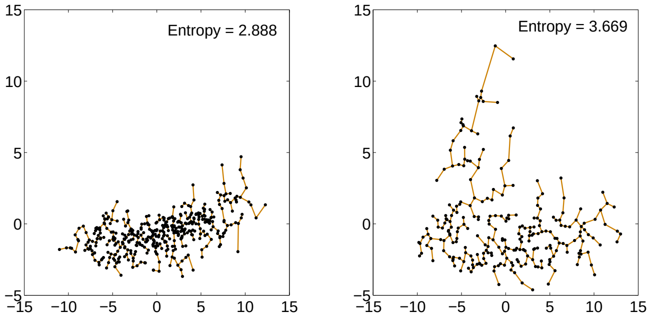

Fig. 2. Comparison of speech feature distributions for two speakers with dysarthria. The speaker on the left has severe dysarthria and the speaker on the right has mild dysarthria.

When comparing the articulation entropy of multiple speakers, we must also consider the differences in length between the resulting speech samples. In order to obtain an unbiased estimate of entropy values, the same number of feature samples should be used to estimate the entropy using the MST approach described above. As a result, we use bootstrap sampling [29]. The sampling process is done multiple times, which allows us to estimate the entropy values for each batch and a final estimate by averaging.

III. Experimental evaluation

We evaluate the approach on two data sets: a database of English-speaking dysarthric patients and a database of Mandarin-speaking dysarthric patients. We provide the Matlab source code for readers who are interested in using this measure[2].

A. *Evaluation on English dysarthric speech*

We use a corpus of English dysarthric speech that contains speech samples from 57 speakers with three types of neurological disorders: 15 speakers with ataxic dysarthria, 16 mixed flaccid-spastic dysarthria, and 26 speakers with hypokinetic dysarthria. Each speaker provided a speech sample consisting of 5 phonetically balanced sentences with an average of 16 syllables in each sentence. The 5 sentences were concatenated into a single utterance - therefore there are a total of 57 speech samples (one per speaker) in this dataset. All speech was sampled at a rate of 16 kHz with 16-bit resolution. Seven second year master's students from the speech and language pathology (SLP) program at Arizona State University were asked to listen to the speech samples

[2] http://www.public.asu.edu/%7Evisar/software/Arti%5FEn.zip

and rate the articulatory precision of the speaker on a scale from 1 (normal) to 7 (severely abnormal). The Evaluator Weighted Estimator (EWE) was used to combine the multiple ratings into a single one by calculating the mean value weighted by individual reliability [30]. The details of this dataset can be found in [31]. In this experiment we evaluate the Pearson correlation between the estimated articulation entropies and the weighted average subjective ratings of all speakers.

1) Evaluation with different feature sets and different window sizes: In Section II-B, four candidate feature sets were briefly described. Here we analyze which feature sets best captures articulatory precision and at what time scale.

After VAD and intensity normalization, the following features were initially extracted from segmented 20*ms* frames (with 10*ms* overlap): 128-D log-spectrum features, 13-D MFCC features, 13-D MelRoot3 features, and 128-D LPC envelope features. Various window lengths ranging from 40*ms* to 200*ms* were also examined. Short-term features were stacked according to different window length settings. Therefore, for a given feature type and a given analysis window, we formed 57 feature matrices - one for each speaker.

The bias and variance of the non-parametric entropy estimator depends on both the sample size and the dimension of the data [32]. Thus, to exclude the influence of these two factors, we used PCA to reduce the dimension of the feature sets and randomly resampled the rows of the data matrix to ensure that each speaker's matrix has the same sample size and same dimension. The dimension reduction is best demonstrated by an example: Suppose the dimension of the stacked MFCC and MelRoot3 features using a 40*ms* long window was 13×4; whereas the dimension of the log-spectrum and LPC features was 128×4. To ensure that all feature sets had the same dimension, we reduced the dimension of the log-spectrum and LPC features from 128×4 to 13×4 using PCA. The sampling process can also be demonstrated with an example: suppose the sample size of the data matrix for each speaker was N_s^i (where i was the speaker index). We randomly sampled each speaker's data matrix using $\left\lfloor 0.9 \times \min_i(N_s^i) \right\rfloor$ samples. Bootstrap sampling was repeated 50 times for each speaker and the articulation entropy was estimated for each batch. The final entropy measure was taken as the average of the individual estimates.

The correlation between estimated entropies and the subjective articulatory precision ratings is shown in Table I. From the table, we can see that the MelRoot3 features with a window length of 160*ms* outperformed the others significantly ($p \ll 0.01$). A correlation value of -0.581 indicates a moderate to strong correlation between the unsupervised articulation entropy metric and the

TABLE I

CORRELATION BETWEEN ARTICULATION ENTROPY AND PERCEPTUAL IMPRESSIONS OF ARTICULATION FOR DIFFERENT FEATURES AND DIFFERENT WINDOW LENGTHS.

	log-spectrum	**MFCC**	**LPC envelope**	**Melroot3**
40*ms*	-0.044	-0.432	-0.343	-0.483
80*ms*	-0.117	-0.469	-0.435	-0.546
120*ms*	-0.162	-0.457	-0.474	-0.568
160*ms*	-0.196	-0.432	-0.505	**-0.581**
200*ms*	-0.196	-0.407	-0.515	-0.564

combined perceptual ratings.

2) Comparison with VSA: For comparison, we also estimated the VSA for each speaker by using the automatic VSA estimation method in [12]. However, this measure shows no correlation (0.08) with the perceptual ratings. Since the VSA relies only on vowel production, the algorithm likely requires more than the five available sentences to form a robust estimate of the vowel space. In addition, the VSA requires robust formant estimation in order to obtain a reliable estimate. For the more severe speakers in the database, this may be difficult to estimate, since most formant estimation methods are tailored to healthy speech.

3) Articulation entropy for voiced/unvoiced speech: We also examined the difference in articulation entropy for voiced (V) and unvoiced (UV) speech samples. A V/UV detection algorithm was used to separate the voiced and unvoiced segments for each speaker. The articulation entropy was calculated on the frames of V and UV speech respectively for each speaker. Random sampling was used to ensure the number of samples for the voiced and unvoiced parts were balanced.

Figure 3 shows the articulation entropies of V/UV segments for all 57 speakers in the dataset. From this figure we can see that: (1) the entropy values of voiced and unvoiced speech samples follow each other, which is consistent with the fact that most dysarthric speakers have both distorted consonants and vowels [10][33]; (2) for some of the speakers, there exists a gap between voiced entropy and unvoiced entropy, which implies that the speaker is able to produce either voiced or unvoiced speech samples more clearly. To show this, two speakers with different profiles were selected: one speaker showing a much larger entropy for voiced than unvoiced speech (Speaker 25), and another showing the opposite (Speaker 38). To listen to the difference,

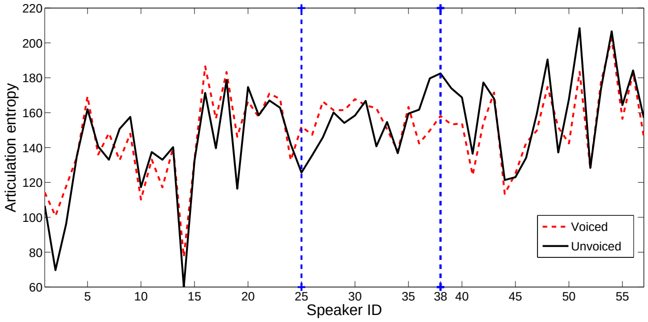

Fig. 3. Articulation entropies for voiced and unvoiced speech samples.

frames detected as voiced were concatenated into a voiced speech stream, and frames detected as unvoiced were concatenated into an unvoiced speech stream. The samples were included as supplemental material. For speaker 25, we can hear that the voiced phonemes (especially vowels) were relatively clear and differentiated, but the unvoiced speech samples were indistinct. For speaker 38, the voiced phonemes were constrained and sound very similar to each other, while the unvoiced phonemes were clear.

B. Evaluation on Mandarin speech

The second dataset we evaluated contains Mandarin speech samples from 10 speakers with PD and an age- and gender-matched healthy control group (7 speakers). Each speaker was recorded in a sound-treated booth while reading the Rainbow passage (translated to Mandarin) [34] into a Shure microphone connected to a Tascam portable digital recorder. In addition, each speaker was recorded in two conditions, a habitual condition and a loud condition. For the habitual condition, the speakers were instructed to speak in their typical manners, while for the loud condition, they were told to speak using a voice twice as loud as their regular talking voice [35]. It has been shown that loud speech can lead to significant intelligibility gains partly because loud speech requires more exaggerated articulator motion [36].

Speech samples in both the habitual and loud conditions were first normalized such that they have the same intensity level. Pauses were removed using the VAD and the articulation entropy was calculated using the MelRoot3 features with a 160*ms* window size. In the left plot of Figure 4 we show the average articulation entropy for both groups and for both conditions in each group. A 2×2 ANOVA test showed that there is a significant difference ($p \ll 0.01$) between the healthy group and the PD group and a significant difference ($p \ll 0.01$) between the habitual and loud

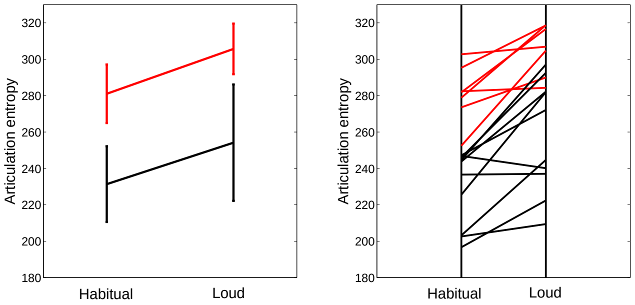

Fig. 4. Articulation entropy for Mandarin speech data in habitual and loud conditions. Red are healthy speakers, and black are PD speakers.

conditions. However, there was no interaction between the two factors. This is consistent with the expectations that (1) the PD group exhibits reduced articulator movement resulting in lower articulatory precision when compared to the healthy group [37] and (2) there is an improvement in the articulatory precision of PD patients speaking in the loud condition [36]. In the right plot of Figure 4 we show the change in entropy for each individual speaker. It can be seen that for all but one speaker, the articulation entropy improved.

IV. Conclusion

Traditional estimation of articulatory precision relies on subjective evaluation or simple objective metrics such as the vowel space area (VSA). In this paper, we propose a more general measure called articulation entropy that serves as a proxy for a speaker's working phonemic inventory. The method extracts features from a continuous speech signal and calculates the entropy of the feature distribution. Compared to the VSA, it is completely unsupervised and does not require any phonetic segmentation. We have shown that the articulation entropy correlates with subjective evaluation of articulatory precision on English dysarthric speech and it captures a clear difference between healthy and PD Mandarin speakers in both habitual and loud conditions.

References

[1] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton, “Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression,” *Journal of psychiatric research*, vol. 27, no. 3, pp. 309–319, 1993.

[2] R. J. Wenke, J. V. Goozee, B. E. Murdoch, and L. L. LaPointe, “Dynamic assessment of articulation during lingual fatigue in myasthenia gravis,” *Journal of Medical Speech-Language Pathology*, vol. 14, no. 1, pp. 13–32, 2006.

[3] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, “Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment,” *Journal of Neurolinguistics*, vol. 17, no. 6, pp. 439–453, 2004.

[4] C. Stewart, L. Winfield, A. Hunt, S. B. Bressman, S. Fahn, A. Blitzer, and M. F. Brin, “Speech dysfunction in early Parkinson’s disease,” *Movement disorders*, vol. 10, no. 5, pp. 562–565, 1995.

[5] V. Berisha, J. Liss, S. Sandoval, R. Utianski, and A. Spanias, “Modeling pathological speech perception from data with similarity labels,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 915–919.

[6] G. Fant, “Speech sounds and features.” 1973.

[7] A. Bladon, “Two-formant models of vowel perception: Shortcomings and enhancement,” *Speech Communication*, vol. 2, no. 4, pp. 305–313, 1983.

[8] G. Weismer, J.-Y. Jeng, J. S. Laures, R. D. Kent, and J. F. Kent, “Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders,” *Folia Phoniatrica et Logopaedica*, vol. 53, no. 1, pp. 1–18, 2000.

[9] R. D. Kent and Y.-J. Kim, “Toward an acoustic typology of motor speech disorders,” *Clinical linguistics & phonetics*, vol. 17, no. 6, pp. 427–445, 2003.

[10] S. Skodda, W. Visser, and U. Schlegel, “Vowel articulation in Parkinson’s disease,” *Journal of Voice*, vol. 25, no. 4, pp. 467–472, 2011.

[11] H.-M. Liu, F.-M. Tsao, and P. K. Kuhl, “The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy,” *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3879–3889, 2005.

[12] S. Sandoval, V. Berisha, R. L. Utianski, J. M. Liss, and A. Spanias, “Automatic assessment of vowel space area,” *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. EL477–EL483, 2013.

[13] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, “Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech,” *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 1, pp. 114–125, 2010.

[14] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, “Acoustic metrics of vowel articulation in Parkinson’s disease: vowel space area (VSA) vs. vowel articulation index (VAI).” in *MAVEBA*, 2011, pp. 173–175.

[15] H. Kim, K. Martin, M. Hasegawa-Johnson, and A. Perlman, “Frequency of consonant articulation errors in dysarthric speech,” *Clinical linguistics & phonetics*, vol. 24, no. 10, pp. 759–770, 2010.

[16] H.-M. Liu, C.-H. Tseng, and F.-M. Tsao, “Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy,” *clinical linguistics & phonetics*, vol. 14, no. 6, pp. 447–464, 2000.

[17] K. Tjaden and G. S. Turner, “Spectral properties of fricatives in amyotrophic lateral sclerosis,” *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 6, pp. 1358–1372, 1997.

[18] C. E. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[19] J. Sohn, N. S. Kim, and W. Sung, “A statistical model-based voice activity detection,” *IEEE signal processing letters*, vol. 6, no. 1, pp. 1–3, 1999.

[20] H. Kuwabara, “Acoustic properties of phonemes in continuous speech for different speaking rate,” in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 4. IEEE, 1996, pp. 2435–2438.

[21] M. Tu, X. Xie, and Y. Jiao, “Towards improving statistical model based voice activity detection,” in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[22] S. Gazor and W. Zhang, “Speech enhancement employing Laplacian-Gaussian mixture,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 896–904, 2005.

[23] J.-H. Chang and N. S. Kim, “Speech enhancement using warped discrete cosine transform,” in *Speech Coding, 2002, IEEE Workshop Proceedings.* IEEE, 2002, pp. 175–177.

[24] V. Berisha and A. O. Hero, “Empirical non-parametric estimation of the fisher information,” *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 988–992, 2015.

[25] V. Berisha, A. Wisler, A. O. Hero, and A. Spanias, “Empirically estimable classification bounds based on a nonparametric divergence measure,” *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 580–591, 2016.

[26] A. O. Hero, B. Ma, O. J. Michel, and J. Gorman, “Applications of entropic spanning graphs,” *IEEE signal processing magazine*, vol. 19, no. 5, pp. 85–95, 2002.

[27] A. Renyi and D. V. d. W. Wahrscheinlichkeitsrechmung, “Berlin, 1966; see also A. Wehrl,” *Rev. Mod. Phys*, vol. 50, p. 221, 1978.

[28] P. E. van der MLJP and J. van den HH, “Dimensionality reduction: A comparative review,” Tilburg, Netherlands: Tilburg Centre for Creative Computing, Tilburg University, Technical Report: 2009-005, Tech. Rep., 2009.

[29] J. J. Higgins, “Introduction to modern nonparametric statistics,” 2003.

[30] M. Grimm, K. Kroschel, E. Mower, and S. Narayanan, “Primitives-based evaluation and estimation of emotions in speech,” *Speech Communication*, vol. 49, no. 10, pp. 787–800, 2007.

[31] J. M. Liss, L. White, S. L. Mattys, K. Lansford, A. J. Lotto, S. M. Spitzer, and J. N. Caviness, “Quantifying speech rhythm abnormalities in the dysarthrias,” *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 5, pp. 1334–1352, 2009.

[32] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, “Nonparametric entropy estimation: An overview,” *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.

[33] J. A. Logemann and H. B. Fisher, “Vocal tract control in parkinson’s disease,” *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, pp. 348–352, 1981.

[34] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, “The nemours database of dysarthric speech,” in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 3. IEEE, 1996, pp. 1962–1965.

[35] K. Tjaden and G. E. Wilding, “Rate and loudness manipulations in dysarthriaacoustic and perceptual findings,” *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 4, pp. 766–783, 2004.

[36] A. T. Neel, “Effects of loud and amplified speech on sentence and word intelligibility in parkinson disease,” *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 4, pp. 1021–1033, 2009.

[37] B. Walsh and A. Smith, “Basic parameters of articulatory movements and acoustics in individuals with parkinson’s disease,” *Movement Disorders*, vol. 27, no. 7, pp. 843–850, 2012.