Does Human Speech Follow Benford's Law?

by

Leo Hsu

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2022 by the
Graduate Supervisory Committee:

Visar Berisha, Chair
Andreas Spanias
Antonia Papandreou-Suppappola

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

Researchers have observed that the frequencies of leading digits in many man-made and naturally occurring datasets follow a logarithmic curve, with digits that start with the number 1 accounting for 30% of all numbers in the dataset and digits that start with the number 9 accounting for 5% of all numbers in the dataset. This phenomenon, known as Benford's Law, is highly repeatable and appears in lists of numbers from electricity bills, stock prices, tax returns, house prices, death rates, lengths of rivers, and naturally occurring images. This paper will demonstrate that human speech spectra also follow Benford's Law. This observation is used to motivate a new set of features that can be efficiently extracted from speech and demonstrate that these features can be used to classify between human speech and synthetic speech.

ACKNOWLEDGEMENTS

I want to extend my gratitude to my thesis mentor Professor Berisha for his guidance in my research and for taking the time to serve as my mentor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Researchers have observed that the frequencies of leading digits in many man-made and naturally occurring datasets follow a logarithmic curve, with digits that start with the number 1 accounting for 30% of numbers and digits that start with the number 9 accounting for 5% of all numbers. This phenomenon is known as *Benford's Law*; for lists that follow Benford's Law, the probability that the first digit in a list of numbers is *d*, *P(d)*, is

$$P(d) = log_{10}\left(1 + \frac{1}{d}\right) \qquad (1)$$

Empirical validation of the law, given a dataset, is straightforward. The first step is to observe the leading digit (i.e., the leftmost digit) for each value in the dataset. The next step is to generate a histogram of the leading digits and to divide each count of leading digits by the total number of values in the dataset. The resulting distribution is the observed empirical Benford distribution, this can be compared with the ideal Benford distribution to see if a dataset follows Benford's Law. While intuitively this is unexpected, this phenomenon has been empirically observed in datasets such as stock prices, tax returns, house prices, death rates, naturally occurring images, etc. Common across these datasets is that they span multiple orders of magnitude; when the data generating process can be modeled as the product of multiple, independent factors, datasets tend to follow Benford's Law [1, 2].

Benford's law is highly repeatable across many datasets that span multiple orders of magnitude. As a result, one of its principal applications is in detection of fraud. For example, accounting auditors analyze tax returns by reviewing the frequency of leading digits to ensure statistical adherence to the distribution in Eqn. (1). If an abnormal deviation from Benford's law is detected, the implication is that the numbers have been artificially modified [3].

Another example of Benford's law used in fraud detection is detection multiple compression of images. In images compressed using JPEG, the Discrete Cosine Transform (DCT) coefficients follow Benford's law [4]. It has been observed that multiple compression of an image changes the DCT coefficients such that they deviate more from Benford's law. If this is observed, the implication is that the image may have been tampered with (e.g., saved multiple times, modified artificially, etc.). Other applications of Benford's Law include detection of scientific fraud [5], e-commerce price distributions [6], macroeconomics [7], etc.

While scientists have shown empirically that many naturally occurring data sets follow Benford's law, to the best of our knowledge, none have analyzed whether human speech follows this phenomenon. The source-filter paradigm models speech as the output of a series of linear time-invariant filters [8]. In the frequency domain, this corresponds to the product of multiple spectra, the same conditions that tend to produce datasets that adhere to Benford's Law [1]; therefore, this observation leads to the hypothesis that speech spectra also follow Benford's Law. In this paper, we evaluate the speech spectra of the several hundred speakers from a phonemically balanced database to determine whether the leading digit follows the distribution in Eqn. (1); we use the results of this analysis to propose a new representation for speech. The principal contributions of this work are as follows:

- We demonstrate empirically that human speech spectra follow Benford's Law
- We propose a new set of speech features based on Benford's Law which can be easily extracted from speech and provide empirical evidence that these features can reliably classify between human speech and synthetic speech

The remainder of this paper is organized as follows. First, we discuss the current applications of Benford's Law, then discuss on what makes a dataset more likely to conform to Benford's Law. Then, we discuss the computational model used to determine whether human speech follows Benford's Law. Next, we propose a new set of features, dubbed the Benford similarity (BenS) features, that can be extracted with little computational burden to characterize the *Benfordness* of speech. In the results section, we confirm that speech follows Benford's Law and empirically demonstrate the utility of BenS features for classifying between human speech and synthetic speech. We end the paper with a discussion of why speech follows Benford's Law and other potential applications of the BenS features.

LITERATURE REVIEW

In this section, we will first discuss the current applications of Benford's Law (Section 2.1), then explain the two explanations for when Benford's Law arises (Section 2.2).

Section 2.1 Current Applications of Benford's Law

As mentioned before, principal uses of Benford's Law is in fraud detection. Two representative examples include auditing account reports and checking for multiple compressions of an image. For the first example, Benford's Law is first modified to find the frequency distribution for the first two digits (rather than just the leading digit). The purpose of utilizing the first two digits is to have a more sensitive test for detection of abnormalities in the dataset [3]. The idealized modified Benford Distribution for the first two digits is:

$$P(d_1 d_2) = log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \tag{2}$$

where d_1d_2 represents the first two digits that range between 11-99, and $P(d_1d_2)$ is the probability the first two digits appear in the dataset. An empirical evaluation of the fit to Eqn. 2 being used can be seen in Figure 1 below.

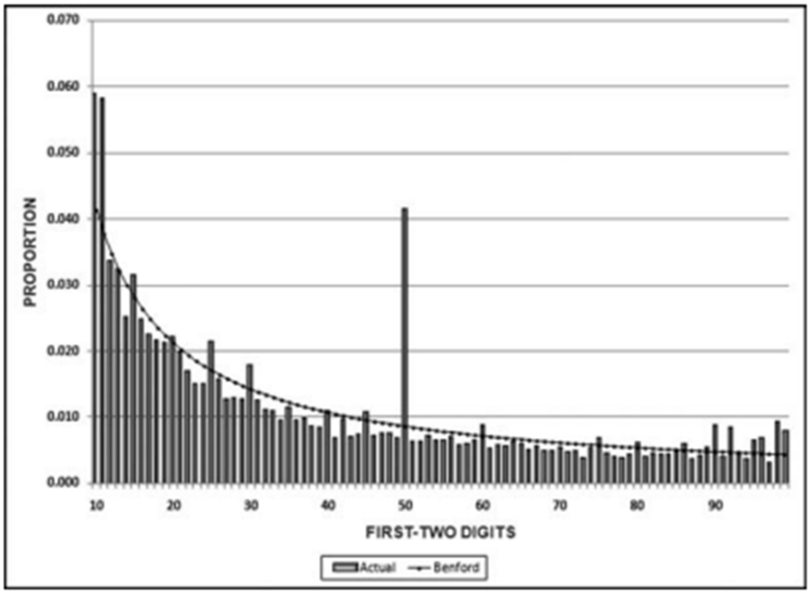


Figure 1: Expense Reports: This plot shows the frequency distribution for the first two digits from an expense report. It is compared to the ideal Distribution for the first two digits. From [3].

According to the figure, there is an unexpectedly large count at around 50, indicating that whoever made the expense report likely artificially modified several numbers starting with 50. This raises a red flag and triggers further investigation to explain the large deviation.

Benford's Law has also been used in image forensics. It has been observed that when an image is compressed into JPEG format, the DCT coefficients that are calculated follow Benford's Law. However, when an image that already went through compression is compressed again, the new DCT coefficients begin to deviate away from Benford's Law, which can be seen in figure 2 below,

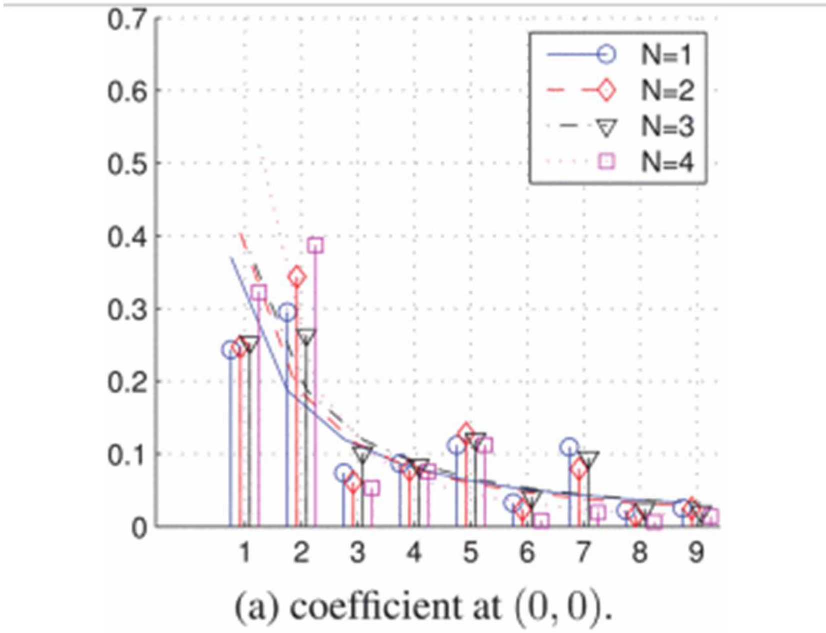


Figure 2: Multiple JPEG Compression: As the numbers of compressions start to increase, the more the DCT coefficients at (0,0) start to deviate from Benford's Law. From [4].

For forensic scientists, noticing any deviation from Benford's Law in DCT coefficients triggers the possibility that someone may have tampered with the image and requires further investigation.

Section 2.2 When does Benford's Law hold?

There are two reasons that give rise to Benford's Law. If a dataset was generated from a product of independent variables, the dataset is more likely to conform to Benford's Law [1]. This can be seen in Table 1.

Data Set	Number of terms in product					
	10 terms		20 terms		50 terms	
	χ^2	Prob	χ^2	Prob	χ^2	Prob
Π(urand(0.0,1.0)	8.21	0.414	-	-	-	-
Π(urand(0.0,0.5)	7.86	0.447	-	-	-	-
Π(urand(0.5,1.5)	15.0	0.0589	6.82	0.55	5.26	0.729
Π(urand(0.9,1.1)	16096	0.0000	10694	0.0000	4930	0.0000
Π(urand(2.0,8.0)	6.99	0.537	6.74	0.565	1.64	0.990
Π(urand(4.0,6.0)	6086	0.0000	2026	0.0000	142	0.0000
Π(urand(4.5,5.5)	17056	0.0000	10947	0.0000	5037	0.0000
Π(urand(4.9,5.1)	108344	0.0000	80260	0.0000	63075	0.0000
Π(norm(0.0,1.0)	2.90	0.935	3.25	0.918	11.0	0.201
Π(norm(5.0,1.0)	707	0.0000	20.4	0.00884	8.87	0.354

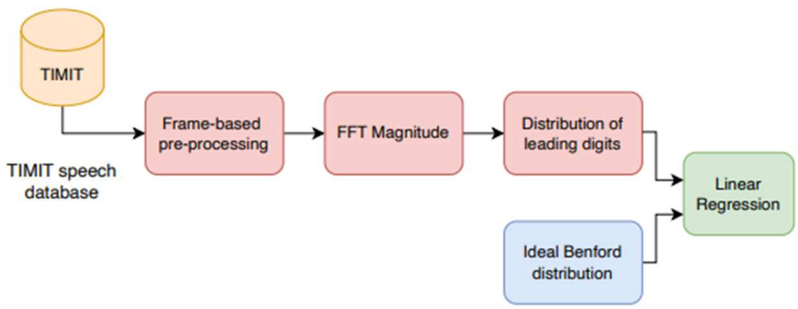
Table 1 Product of Independent Variables: Shows the Chi-Square test between the observed Benford distribution from a dataset generated from a product of independent variables and the ideal Benford distribution. From [1].

The table above shows 3 different experiments to determine whether datasets generated from a product of multiple independent variables are more likely to conform to Benford's law. The first experiment was to generate a dataset where we first draw 10 numbers from a random distribution, then multiply all 10 numbers together. The author repeats this process 10,000 times to generate a new dataset, then the Benford distribution is calculated using the steps mentioned earlier. Once the observed Benford distribution is calculated, a Chi-Square Test is measured between the observed and ideal Benford distribution. As seen in Table 1, there are multiple instances where the *p*-values are greater than the 5% significance level, meaning the null hypothesis can be accepted, which states that there are no differences between the observed and ideal distribution. The second and third experiment are very similar to the first experiment, the only difference is rather than multiplying 10 numbers together, 20 numbers are multiplied together for the second experiment, and 50 numbers for the third experiment. This provides evidence that when the data generation process consists of products of multiple independent random variables, the resulting dataset follows Benford's Law. This product of multiple values can also lead to datasets that span across multiple orders in magnitude. This can be observed in expense reports and tax returns, where the values can range from \$100-\$100,000.

METHODS

In this section we provide an overview of the methodology to evaluate whether speech follows Benford's Law (Section 3.1) and propose a new set of speech features based on Benford's Law (Section 3.2).

Section 3.1 Evaluating Whether Speech Follows Benford's Law



Compare Benford with Speech Block Diagram

Figure 3: A block diagram of the approach used to evaluate whether speech follow's Benford's Law.

In Fig.3 we provide a block diagram of the computational methodology used to determine whether speech follows Benford's Law. The TIMIT database is used for analysis [9]. The TIMIT train set contains audio recordings from 326 male and 136 female speakers; in this corpus, each speaker produced 10 sentences - 2 sentences common to all speakers and 8 non-overlapping sentences. This dataset was used to demonstrate whether the speech spectrum follows Benford's Law by comparing the probabilities in Eqn. (1) with estimates derived from the frequency of leading digits in the speech spectrum. Prior to analyzing the audio, dithering was used to randomize quantization errors that appear in the audio. That is, we intentionally add a small amount of noise to remove trails of zeros that were the result of quantization in the data. The noise was Gaussian distribution with mean 0 and variance determined by dividing the maximum value of the speech signal by 1000. This ensured that the additive noise had no perceptual impact on speech quality.

Our working hypothesis was that Benford's Law would be apparent in the spectrum of speech signals. To evaluate the hypothesis, we use frame-based processing to form the short-time Fourier transform of the signal. Each speech sample was split into 25ms frames with 10ms overlap. For each frame, the DC component of the speech signal was removed by subtracting the mean from the speech signal frame. Next, we calculate the FFT magnitude of the resulting vector to acquire the amplitude of the speech frequencies for each frame in a signal. We remove the first element from consideration since it is zero (after de-meaning) and the resulting vector is normalized by the smallest value so that all values are greater than 1. It is also to make the values in the FFT magnitude to span across multiple orders of magnitude since it is known datasets are more likely to conform to Benford’s Law as mentioned before. The leading digits of the resulting vector are extracted to estimate their respective probabilities via a frequency histogram. This process is repeated for all frames in a speech signal, then across all audio samples from all speakers in the TIMIT database. We calculate the probability distribution for each speaker by normalizing the frequency histogram by the total number of frames for that speaker, then find the average across all speakers. We compare the resulting average with the ideal Benford distribution using a linear regression model.

Section 3.2 Benford Similarity (BenS) Features

As we will present in Section 4, when averaged over multiple frames, multiple speech samples, and multiple speakers, speech follows Benford's Law. This finding motivates a new feature set for characterizing the *Benfordness* of a given speech sample on frame-by-frame basis. Briefly, this feature set estimates the leading digits' probabilities for a single speech frame, compares the distribution for a frame to the ideal Benford distribution, then repeats the process for all frames in a sample and characterizes the similarity across the sample using several statistics. A high-level block diagram of the feature extraction steps is shown in Fig. 4.

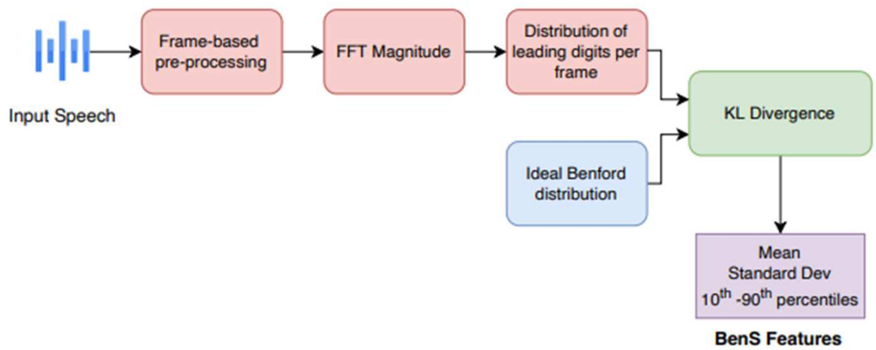


Figure 4: KL Block Diagram – This block diagram shows the approach used to extract the BenS Features.

To extract the speech features, a speech sample is analyzed using frame-based approach with 25ms frames and 10ms overlap. Next, we find the FFT magnitude of each frame and extract the leading digits from the amplitude at each frequency component. The leading digits are used to estimate the empirical Benford distribution, $\hat{P}(d)$,which is compared against the ideal Benford Distribution, *P(d)*, using the Kullback-Leibler (KL) Divergence,

$$D_{KL}\left(P\middle|\middle|\hat{P}\right) = \sum_{d=1}^{9} P(d)log_{10}\left(\frac{P(d)}{\hat{P}(d)}\right) \quad (3)$$

The KL divergence evaluates the similarity between the two probability mass functions; a small KL Divergence implies that the distributions are similar, whereas a large KL divergence implies that the distributions are very different. Lastly, if the KL divergence equals to 1, the two distributions are identical.

For each frame, we obtain a single feature, the KL divergence between the true and the empirical distributions. This process is then repeated across all frames in a sample. For some frames, the empirical estimate of the less frequent numbers in the Benford distribution are 0, which makes it impossible to estimate the KL Divergence (since it goes to infinity). As a result, we remove all frames which have *0* estimates for any of the empirical probabilities. The distribution of the remaining KL divergence values is characterized using 11 statistics; these form the final feature set extracted at the sample level. These statistics include: (1) mean of the KL, (2) standard deviation of the KL, (3)-(11) the 10^{th}-90^{th} percentiles of the KL. We name this feature set the Benford Similarity (BenS) Features. The purpose of these statistics is to demonstrate the average and the spread of the KL values across all the frames in the speech signal. As frequencies change over time in the signal, so will the KL values and it is important to see how KL values change as frequencies in the signal change.

RESULTS

In this section, we present the results of our analysis. In section 4.1, we provide evidence that speech follows Benford's Law. In section 4.2, we empirically evaluate the utility of the BenS Features to distinguish between human speech and synthetic speech.

Section 4.1 Does Human Speech Follow Benford's Law?

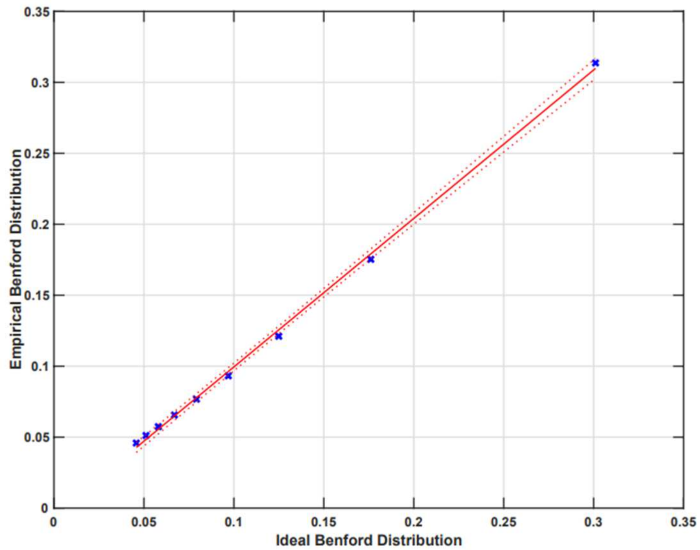


Figure 5 Benford Speech Linear Plot: This figure shows the agreement between the empirical and ideal Benford distribution.

In Fig. 5, we show a scatterplot comparing the probabilities from the ideal Benford distribution with the empirical probabilities estimated by average across all speakers in the TIMIT corpus, per section 3.1. It's obvious from visual inspection of this figure that there is strong agreement between the empirical and the ideal distributions. A linear regression between empirical (E) and ideal (I) Benford probabilities yields the equation:

$$E = 1.046\, I - 0.005 \tag{4}$$

A slope near 1 and an intercept near 0 provides further evidence that speech spectra follow Benford's Law. The association is confirmed statistically as the model has $R^2 = 0.99$.

Section 4.2 Using BenS Features to Detect Synthetic Speech

Dataset generation: Now that there is empirical evidence that human speech follows Benford's Law on average, and with deepfake technology becoming increasingly more sophisticated, new techniques are required to distinguish human speech from synthetic speech. To that end, we evaluate whether BenS features can be used to classify between real speech samples and synthetic samples.

We consider 20 different US-English voices from male speakers and female speakers from Google's text-to-speech API. The Google API includes 10 TTS voices based on a pre-trained Wavenet model with a high mean opinion score (MOS) and 10 standard voices based on other models [10]. We compare speech from these synthetic voices to 20 speakers from TIMIT. We match the gender and the content of the spoken text across the two groups by selecting 20 TIMIT speakers from the TIMIT DR2 test set. We used the 10 sentences spoken by each of the 20 human TIMIT speakers to generate matched samples from the 20 TTS voices. We extracted the BenS Features from each of the 10 sentences spoken by 20 human speakers and 20 synthetic speakers. This results in two data sets, each consisting of 200 samples (20 speakers * 10 samples/speaker); from each sample we extract the 11 BenS features.

Classification experiment: The two datasets were used to train a classifier to distinguish between human speech and synthetic speech. We evaluate the model using leave-one-speaker-out (LOSO) cross-validation. That is, we remove one speaker from the training set and train the model using the data from the remaining speakers. After training the model, the data from the removed speaker is used to evaluate the accuracy of the model. This is repeated across all speakers, one by one, and the resulting estimate of accuracy is considered the out-of-sample estimate of the model's accuracy. We considered a linear support vector machine (SVM), a decision tree, and several higher-order polynomials-kernel SVMs. While all models performed reasonably well, the best performing model was the quadratic SVM (QSVM) with an accuracy of 91.5 percent and a misclassification rate of 8.5 percent. Table 2 shows the confusion matrix of the QSVM model.

Confusion Matrix

	Predicted Classes	
True classes	Human	Synthetic
Human	181	19
Synthetic	15	185

Table 2: The confusion matrix for the classifier that uses the BenS Features to distinguish between synthetic speech and human speech.

It's clear from the results that the BenS features separate human speech from synthetic speech. That is, there is a difference in the distribution of BenS features between human speech and speech generated by existing TTS systems. We can visualize this difference in distributions between human speech and synthetic speech via a scatterplot of the first two BenS features, the mean and standard deviation of the KL divergence values estimated for a given sample. In Fig. 6, we plot these two features for the 200 samples from the 20 human speakers (blue) and the 200 samples from the 20 synthetic speakers (red). The figure clearly shows separability between the two groups.

BenS feature normalization: We use the TIMIT training data used to confirm that speech follows Benford's law to normalize the features extracted from the 20-test human

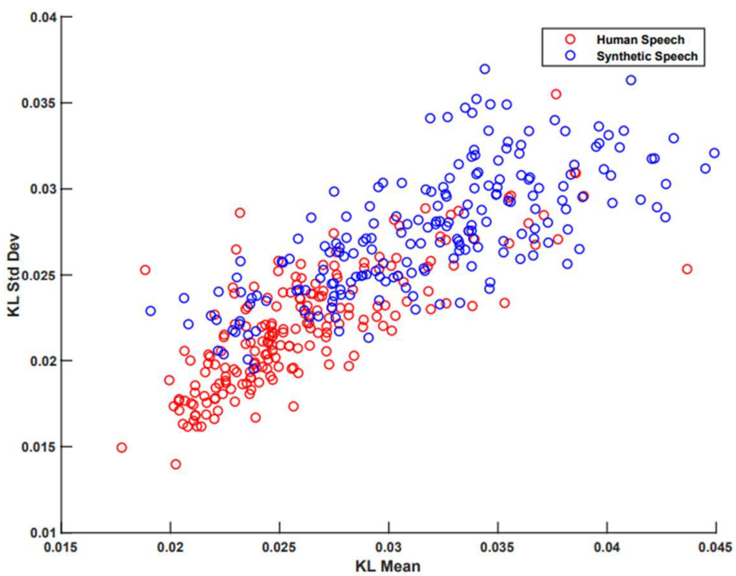


Figure 6 Scatter Plot for KL Divergence: A scatter plot of two BenS Features, the mean of the KL Divergences, and the standard deviation of the KL Divergence.

speakers and the 20 synthetic speakers. That is, the mean and standard deviation for each of the 11 BenS features from the TIMIT training set was used to *z*-score the features from the data used in the classification experiment,

$$z_i = \frac{x_i - \mu_i}{\sigma_i}, \quad (5)$$

where x_i is the i^{th} BenS feature, μ_i and σ_i are the mean and standard deviation of that feature (estimated from the TIMIT training set), and z_i is the normalized version of that feature. Figure 4 also shows that the average KL values for the human speech samples are closer to zero than the KL values from the synthetic speech, indicating that human speech is more likely to follow Benford's law than synthetic speech. The plot also shows the difference in variability between the two speech samples. The synthetic speech samples have a higher variability compared to human speech which can be seen by observing the Standard Deviation feature between human speech and synthetic speech.

Z-scores

Feature	Human Speech	Synthetic Speech
Mean	0.859	1.861
Standard Deviation	0.814	1.812
10th Percentile	0.784	1.409
20th Percentile	0.814	1.379
30th Percentile	0.814	1.401
40th Percentile	0.847	1.423
50th Percentile	0.850	1.579
60th Percentile	0.856	1.915
70th Percentile	0.874	2.320
80th Percentile	0.892	2.358
90th Percentile	0.874	1.966

Table 3: z-scores for the BenS Features for human and synthetic speech. The means and standard deviations for the normalization were obtained from the TIMIT train set.

Table 3 shows the *z*-scores for each feature extracted from human and synthetic speakers. It is clear there are differences between the features for human and synthetic speech. The human speech samples extracted from the 20 human speakers have *z*-scores within approximately 1 standard deviation of the feature distributions estimated from the larger TIMIT training set. In contrast, the synthetic speech has *z*-scores that exceed 1 standard deviation from the mean. On the one hand, this finding makes sense as we would expect that synthetic speech samples are distinct from human speech given the results of the classification analysis. On the other hand, the *z*-scores for the 20 human speakers are not zero; rather, they are positively biased to approximately ~0.8 standard deviations from the mean. We posit that this bias is a result of a mismatch in the data distributions between the complete TIMIT training set and the small subset of speakers we used from the TIMIT test set. Our TIMIT test set contained speakers from only one dialect region (DR2), whereas the TIMIT training set contains data from all dialect regions. We controlled for the dialect region in our test set to reduce variability for our classification analysis given the small sample size. The sample size was limited given the small number of synthetic voices we had available.

DISCUSSION

The evidence in Fig. 3 and the results of the linear regression analysis suggest that speech spectra follow Benford's Law on average. However, variability was not accounted for when comparing the human speech spectra with the Benford distribution. Meaning if each speaker’s Benford distribution were compared with the ideal distribution, there would be some instances where the linear regression would display the differences between the human speech Benford distribution and the ideal distribution. The R^2 value would also decrease as well. Even though variability was not accounted for when comparing human speech with Benford’s Law, the variability between human speech spectra and synthetic speech spectra is different as shown in Figure 4. Indicating that the variability is a good feature to use to help classify between human and synthetic speech.

The result of the average human speech Benford distribution following Benford’s law is not surprising when we consider existing models of the human speech production mechanism. Benford’s Law is apparent in datasets where the data generating process is the product of multiple independent factors [1]. Source-filter theory models speech as the output of a sequence linear time-invariant systems at the frame level [8]. That is, in the frequency domain, we model the speech spectrum $S(\omega)$ as,

$$S(\omega) = E(\omega)\, V(\omega)\, R(\omega), \qquad (6)$$

where $E(\omega)$is the glottal excitation signal in the frequency domain, $V(\omega)$ is the spectrum of the impulse response of the vocal tract filter, and $R(\omega)$, is the spectrum of the impulse response of a filter that models the radiation characteristics at the mouth. The equation above clearly models speech as the product of multiple factors. While $E(\omega)$, $V(\omega)$ and $R(\omega)$, are not completely independent, most instantiations of the source-filter model make the simplifying assumption that they are. Adherence to Benford's law provides additional evidence that this independence assumption is reasonable as Benford's Law holds for datasets generated via a product of multiple *independent* factors [1].

We further demonstrated that the BenS features are useful to classify between human and synthetic speech. The distribution of KL divergence values over a sample – as measured by the BenS features – were markedly different for human speech and synthetic speech. The differences in the BenS features are perhaps not surprising when we consider differences in the data generating process between human speech and synthetic speech. Human speech is modeled as the output of Eqn. (5); however, the most sophisticated synthetic voices used in our database were produced by neural networks with complex non-linear structures. It's possible that the BenS features capture these fundamental differences in the underlying mechanism of production; however, evaluating this hypothesis requires a deeper study of synthetically generated speech.

While in this paper we demonstrated that the features are useful for detection of synthetic speech, other applications of BenS features may be possible. For example, for speech produced by patients with certain clinical conditions, the independence assumption between $E(\omega)$ and $S(\omega)$ is more difficult to justify. In neurological conditions where velopharyngeal control is impacted (e.g., amyotrophic lateral sclerosis [11]), air escapes through the nasal cavity during speech (changing $V(\omega)$), therefore the speaker may actively attempt to increase the loudness of their speech (thereby changing $E(\omega)$). This clearly makes $E(\omega)$ and $V(\omega)$ dependent on each other. We conjecture that this would lead to changes in the BenS representation. As a result, follow-on work should evaluate the utility of this representation for detecting early changes in clinical conditions via speech. If validated, the computational simplicity of the feature set would make it ideal for persistent on-device tracking.

Another possible future application for BenS Features is to classify between different musical instruments. Each instrument has their own sound, just like how every human have their own voice. Since each instrument has their own unique spectra, future work would be to experiment whether the BenS features for each instrument are different and if it is possible to classify between different instruments using BenS Features rather using current methods involving neural networks which has more computations than the computations required for BenS Features.

CONCLUSION

We evaluated the distribution of leading digits in the speech spectrum (extracted at the frame level) and showed that it follows Benford's law on average. Variability of human speech Benford distribution was not accounted for, but its variability compared to the synthetic speech Benford distribution shows a significance difference where a classifier was able to differentiate the two 91.5 % of the time. One possible explanation for adherence to Benford's law comes from existing models of the speech production mechanism. Human speech can be modeled as the product of multiple independent signals in the frequency domain, and it's known that datasets that can be modeled in this way will tend to follow Benford's Law. Based on this observation we proposed a new feature set, dubbed the Benford similarity (BenS) features, for characterizing the *Benfordness* of speech at the level of an utterance using the KL divergence between the empirically estimated Benford distribution and the ideal Benford distribution. Classification results showed that this feature set can distinguish between human speech and synthetic speech with accuracy exceeding 90%. It's important to note that the classification experiment was limited as the dataset consisted of only 20 speakers from the TIMIT test set. This was done to match the human data to the unique synthetic voices, of which we only had 20. Future work would involve utilizing different APIs that have more different synthetic voices available so the dataset would increase and be able to train the classifier and determine whether the performance increases or decreases. Future work will focus on analyzing the utility of BenS Features in other contexts such as for musical instruments and for patients with clinical conditions. As mentioned before, patients with ALS would have less control of their velopharyngeal valve, which lets some air escape through their nose as they speak, which implicates the vocal tract and the patient would have to speaker louder to be heard, which would affect the glottal signal as mentioned. Those patients would have a lower frequency range in their speech , indicating it should not follow Benford's Law since a dataset should have values that span across multiple orders of magnitude as mentioned before. Future work would involve observing the speech spectra and compare it with the ideal Benford distribution and determine whether the spectra follows or deviates from Benford's law.

Future work will also involve utilizing BenS Features on musical instruments. There are currently multiple different methods on instrument recognition, so future work would see if BenS Features could be a new method that require less computations and less features and still be able to accurately classify between different instruments.

REFERENCES

[1] P. D. Scott and M. Falsi, “CSM-349-Benford’s Law: An Empirical Investigation and a Novel Explanation,” *Technical Report*, CSM-349, University of Essex, Colchester, 2001.

[2] F. Benford, “The Law of Anomalous Numbers,” in *Proc American Philosophical Society*, 78(4), pp 551-772. 1938.

[3] J.M. Nigrini, “Benford’s Law: Applications for forensic accounting, auditing, and fraud detection,” *John Wiley & Sons*, vol. 586, 1989.

[4] S. Milani, M. Tagliasacchi, and S. Tubaro, “Discriminating multiple JPEG compression using first digit features,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2253-2256, 2012.

[5] S. Hüllemann, G. Schüpfer, and J. Mauch, “Application of Benford’s Law: a valuable tool for detecting scientific papers with fabricated data?,” in *Der Anaesthesist* 66.10, 2017, pp. 795–805.

[6] E. D. Giles, “Benford’s Law and naturally occurring prices in certain ebay auctions,” in *Appled Economics Letters 14.3*, 2007, pp. 157–161.

[7] Nye, John, and Charles Moul, “The political economy of numbers: on the application of Benford’s Law to international macroeconomic statistics,” in *The BE Journal of Macroeconomics 7.1*, 2007.

[8] Quatieri, T. F. “Discrete-time speech signal processing: principles and practice.” Pearson Education India, 2006.

[9] J. S. Garofolo, et. al, ”TIMIT Acoustic-Phonetic Continuous Speech Corpus”, LDC93S1, Web Download, Philadelphia: Linguistic Data Consortium, 1993.

[10] Google TTS voices: https://cloud.google.com/text-to-speech/docs/voices, Accessed March 2022.

[11] J. R. Green, et. al, ”Bulbar and speech motor assessment in ALS: Challenges and future directions”, in Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 14:7-8, pp. 494-500, 2013.