Laughter signature, a new approach to gender recognition

Comfort O. Folorunso | Oluwatoyin P. Popoola | Olumuyiwa S. Asaolu

Abstract
Gender Recognition (GR) is the process of identifying gender difference by extracting and evaluating features from data existing as images, video, audio, text, and other signals. The task of GR has been achieved using face, keystroke, gait, and speech features. “Laughter” with its intrinsic usefulness for this task is yet to be explored. Laughter is an important paralinguistic feature in human communication. It can change the meaning of speech when triggered by some form of arousal or amusement. Although, laughter can be “acted,” but the human natural laughter is a spontaneous reflex response, which reasonably embeds some characteristic peculiarities of the individuals. The human brain has capacity to make this distinction. In this study, spontaneous laughter bouts of 123 volunteers (41 females and 82 males) were recorded. The Dynamic-Average of the Mel frequency Cepstral Coefficient (DA-MFCC) were generated and trained using two conventional and effective machine learning algorithms that have been employed in gender identification. These algorithms are Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) classifiers. Overall accuracies of 87.65% and 86.91% were obtained with GMM and SVM respectively. Therefore, indicating the possibility of using laughter characteristics as signatures for distinguishing between male and female genders. For both classifiers, the use of DA-MFCC reasonably reduced training time. Some of the potential areas of applications of GR include security, health care, marketing, human machine interaction toward enhanced emotion recognition, automatic speaker recognition and forensics.

KEYWORDS
Gaussian mixture model, gender recognition, laughter signature, Mel frequency cepstral coefficient, support vector machine

1 INTRODUCTION

Laughter is a vital paralinguistic event in human communication. The automatic detection of laughter occurrences in human speech can enhance automatic speaker recognition (ASR) systems, automatic speech recognition system, identify humorous content in video clips, and detect speaker’s emotion. Therefore, incorporating laughter detection in automatic speech recognition systems can reduce word error rate by recognizing non-speech sounds, thereby improving the system.1
This work proposes the use of the audio laughter only, and excludes the visual cues for gender recognition (GR) because laughter can be heard from a distance or in the dark, and be perceived or recorded without imagery.

Gender is a multifaceted concept, which plays an imperative role as a social construct and essential form of an individual’s personality. This work is based on the traditional gender profiles of “male” and “female,” it does not cover hermaphrodite and transgender categories. Research in this area has progressively disclosed its complex form, which shows a lot more variability. The ability to differentiate male from female is crucial to the survival of the human species (e.g., in cases of reproduction, education, safety, and security such as crimes scenes investigation). Moreover, nature offers a myriad of visual cues through which genders can be differentiated. These vary from secondary sensual features to slight behavioral cues including gait. Consequently, some of the core visual biometric technologies initially built for person identification have demonstrated the ability to differentiate between genders, with reliable precision. GR can, for example, be achieved using low frequency data from the outline of a human face; kinematic data from gait analysis; skin texture; keystroke; voice; and speech.

Recently some group of researchers worked on Parkinson’s disease detection and GR using the deep network including pooling and feature extraction methods, to improve on the existing methods on GR. For instance, Tuncer and Dogan and Yaman et al worked on Parkinson’s disease and gender detection. They introduced different pooling methods including the minimum, maximum, mean, and median (M4) pooling methods, the octopus-based feature network and a statistical pooling method for GR. Findings from their works revealed that the pooling methods outperformed convolution methods in existence. Also, Tuncer et al and Ertam reported a 1-dimensional local binary pattern feature extraction network (1D-LBPNet) for feature extraction and deeper Long Short-Term Memory network with some conventional machine learning algorithms for GR. Their results showed that these methods have a high success rate when compared with existing GR methods.

GR is usually a binary classification task. In voice processing, it is a technique that is often used to establish the gender group of a speaker through the analysis of speech cues. Currently, there are different approaches, which have been documented on voice-based GR. This voice-based GR plays a crucial role in the context of various usages such as voice disguise recognition, forensic surveillance, intelligent human-machine interaction, emotion recognition, mobile healthcare system, and Interactive Voice Response systems. It is also useful during the pre-processing stage in some speech-based roles including speaker verification and diarisation. Studies have shown that GR helps to improve accuracy in these different tasks. Similarly, it can be used in speaker characterization by predicting emotion or age either concurrently or conservatively. Another usage of gender-based person identification is client interaction system via voice query, which helps client in decision making. Similarly, GR is beneficial in mobile healthcare systems, particularly in the instance of massaging services (reservation system) that demands a customer to be massaged by the appropriate person within the business atmosphere.

Some of the studies on GR used voice frequency and pitch to classify speakers. This has made scholars to recognize that female voice has higher frequency than male voice. Nevertheless, pitch and frequency alone are not adequate to differentiate gender. A good GR system must be able to function well in unfavorable states such as background noise, session or channel inconsistency, recording environment, and language variation among others. These parameters, therefore, make GR from voice samples a demanding task. In other studies, speech processing used to process voice samples, has through the years given rise to voice biometrics, emotion recognition, and automatic identification of age as well as gender. Speaker’s gender and age classification is one of the most demanding tasks in the field of speech processing. This is because of background noise, duration of speech, text-dependent or text-independent system design, and intonation variance due to different speakers. These studies have shown that GR improved client satisfaction and relationships among other essential causes and vital in ASR systems, since gender-dependent systems perform well compared to gender-independent systems. Thus, various approaches are continually being proposed by researchers, in the quest to enhance ASR systems.

In recent times, several studies have been carried out on age, gender, and emotion recognition using different techniques. For instance, the raw face images, both local and low-level features from the face image and face images combined with voice have been used for GR studies. Ugail and Al-Dahoud as well as Dantcheva and Bremond studied the efficiency of gender identification from dynamic smile parameters. Their works used Support Vector Machine (SVM), K-Nearest Neighbors, Adaboost, and Bagged Tree as classifiers. They established the existence of strong indicators of gender dimorphism. From the foregoing, studies have shown that the dynamics of posed smiles carries significant cues on gender, and it is more discriminant for people less than 20 years, while the facial appearance features are more discriminant for people above 20 years of age. However, there is a significant difference between smile and laughter. The former is based on imagery while the latter is rooted in audio. According to Nagata and Mori, there is a gradual change...
from smiling to pure laughter because of the transition from speech-smile to speech-laugh to pure laughter. Each of which has different acoustic characteristics and communicative functions.

Studies have shown that audio features such as the long-term structure of the audio spectrum, its derivatives and transformations, and Perceptual Linear Predictive, are useful for gender classification and emotion recognition. Acoustic and prosodic features of speech are other techniques that have been used for this classification and recognition. Also, Gaussian Mixture Model (GMM), SVM, Artificial Neural Network, and Deep Neural Network are among the “state-of-the-art” tools, which are used for feature extraction and gender classification. In another study, Livieris et al. used an ensemble semi-supervised self-labeled algorithm, known as iCST-voting, in solving the GR problem. This technique uses ensembles to combine co-training, self-training, and tri-training using an ensemble as base learner. Their findings demonstrated the efficiency of iCST ensemble algorithm over other “state-of-the-art” self-labeled algorithms. It has also been shown recently that recognizing speaker’s age and gender improved the efficiency of speaker verification and identification systems including Human Computer Interaction. These systems can be customized, based on speaker’s voice signal to manage discussion and to enhance the stage of client satisfaction.

Human voice can produce both verbal (spoken words or speech) and non-verbal (coughing, crying, hissing, and laughing) sounds. Researchers over the years have concentrated on speech for voice-based biometric analysis. This work, however, explores another plausible human characteristic that is naturally occurring and can be less mimicked, which can be used as a biometric for gender identification. Laughter has been observed to be a very common vocal sound and like human voice, appears to have consistent style or patterns peculiar to individuals and may be useful in characterizing gender. This stimulus-triggered natural expression is often not easily suppressed unlike speech. It can be used in situations where speech is blurred, inaudible or unavailable. These characteristics form the rational for investigating laughter as a potential biometric for gender classification. Speech and laughter may be similar in frequencies and syllable but there are still some acoustic and spectral features differences between laughter and speech as discussed by Nagata and Mori as well as Truong and Leeuwen. According to Dumpala et al., laughter is highly variable as compared to speech; because the glottal configuration of laughter is different from speech due to high subglottal pressure. Speech is controlled while laughter can be voluntary (acted) as well as involuntary (spontaneous). Spontaneous laughter is hardly controllable once it is aroused. Whereas speech/voice can be mimicked perfectly, laughter can hardly be mimicked. Therefore, laughter when established may be more beneficial for forensic use. The differences between laughter and speech features can be exploited using the fundamental frequency also known as pitch and the formants frequencies, which are higher in laughter than speech. In laughter, there is an increase in the airflow through the vocal tract that results in faster vibration of the vocal folds. The articulate expression of laughter changes across sex, individuals, and settings as different individuals have their favorite subset of phones for laughing.

Ruch et al. pointed out that there are biometric traits in laughter like in fingerprints and confirmed that this field has not been exploited. Thus, the aim and major contribution of this study is to investigate the use of laughter signature for GR using audio laughter.

2  |  METHODOLOGY

2.1  |  Theoretical background

Human perception cannot measure minute changes in frequency like specific-engineered microphones do. However, human perception can more easily distinguish “pitch.” In other words, machines detect frequency while human beings perceive pitch. Pitch is the perceptual equivalent of frequency. Human perception of pitch follows a linear scale for up to 1000 Hz but a logarithmic scale afterwards because the ability to distinguish pitch becomes fuzzier as the frequency increases. In order to cater for this non-linearity, an approximation of human perception of pitch is represented as Mel Frequency Cepstral Coefficients (MFCC). These coefficients are computed in the frequency domain on a logarithmic “Mel Scale” and used as a feature in sound and speech processing systems.

Sounds produced by humans are filtered by the shape of the vocal tract and some articulators including the tongue, teeth, and related organs. This shape defines what sound is produced. If this shape can be accurately determined, the speech sounds that are produced can be correctly distinguished and represented. The vocal tract structure reveals itself in the envelope of the short time power spectrum, which MFCC effectively represents.

The mathematical transformation of a frequency from the Hertz to the Mel scale is given in Equation (1), while Equation (2) is the inverse of the Mel scale transformation described in Equation (1).
\[
\text{mel} (f) = 2595 \times \log_{10} \left( 1 + \frac{f}{700} \right) \quad \text{(1)}
\]

\[
f = 700 \left( 10^{\frac{\text{mel}}{2595}} - 1 \right) \quad \text{(2)}
\]

In this study, GMM and SVM algorithms are used as classifiers for training, testing, and validation. These classifiers are easy to implement and highly efficient in solving classification problems. SVM is widely used within the Machine Learning community. GMM on the other hand is a probabilistic distribution model where the main distribution is assumed to be Gaussian in nature such that the distribution has a single peak. GMMs can also model distributions with multiple peaks by adding a number of Gaussians together. Employing an adequate number of Gaussians, varying their means, covariance, and weights makes it possible to approximate any continuous density to a random accuracy. It is the most frequently used distribution system for modeling speaker and gender features in person and GR systems because speech features are usually assumed to be normally distributed. The ease of learning ability, boot strapping from flat data faster computation, and compatibility with frame-level features makes it a choice modeling tool. GMM has also been proven to be very efficient and effective in text-independent classification system where the prior knowledge of what the speaker will say next is not known. It approximates the mean and covariance of each individual speaker based on a feature vector. However, in a situation where one Gaussian is not sufficient to describe a density, mixture models are implemented.

A mixture of Gaussians can be described in Equation (3).

\[
p(x) = \sum_{k=1}^{K} \Pi_k N(x | \mu_k, \Sigma_k)
\]

where \( x \) is a \( d \)-dimensional vector, \( \Pi_k \) is the weight of the \( k \)th Gaussian component, \( \mu_k \) is the \( d \)-dimensional vector of means for the \( k \)th Gaussian component and \( \Sigma_k \) is a \( d \times d \) covariance matrix for the \( k \)th Gaussian component, and \( N \) is a \( d \)-dimensional Gaussian of the form presented in Equation (4).

\[
b_k(x) = N(x | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp \left( -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \right)
\]

where \(|\Sigma|\) is the determinant of \( \Sigma \).

By using the density function \( b_k(x) \) and the mixture weights \( w_i \), the GMM can be described as a weighted sum of \( M \) component densities as shown in Equation (5).

\[
p(x | \lambda) = \sum_{i=1}^{M} w_i b_i(x)
\]

where \( x \) is a \( D \)-dimension random vector;

The parameter \( w_i \) are the mixing coefficients, which satisfies conditions \( 0 \leq w_i \leq 1 \) and \( \sum_{i=1}^{M} w_i = 1 \).

Each Gaussian density \( b_i(x) \) is called a component density. Each component density is described by the parameters \( \mu_i \) and \( \Sigma_i \).

Each speaker is characterized by a mixture of mean, variance, and weights as presented in Equation (6).

\[
\lambda_i = \left\{ w_i, \mu_i, \Sigma_i \right\} ; i = 1 \ldots M
\]

SVM is used for binary classification. It locates a hyperplane in an \( N \)-dimensional space that specifically classifies the points, with \( N \) as the number of features. The hyperplane must have a maximum distance between the data points of the two classes. This distance between the decision boundary and the closest data point established the margin of the classifier. This is necessary in order to give some support and ensure an accurate classification when a new data point is presented for classification. Support vectors are data points nearer to the hyperplane. Conversely, multi-class SVMs are made up of several binary SVMs fused together. During the training phase, the SVM takes a matrix of features as input, and classifies each sample as a member of a particular class (ie, positive outcome) or not (negative outcome). Each data
sample in the feature matrix is taken as a row in a high dimensional feature space, where the dimension is equivalent to a number of attributes. The SVM establishes the most appropriate hyperplane, which classifies each positive and negative training sample.

The decision boundary of a linear classifier is defined in Equation (7).\(^ {54}\)

\[
w^T x + b = 0
\]  
\[(7)\]

where “\(w\)” is weight vector and “\(b\)” is bias term.

For a given corpus and decision hyperplane, the functional margin of the \(l\)th sample \(x_i\) with respect to a hyperplane \((w, b)\) is defined in Equation (8).\(^ {54}\)

\[
y_i = y_i (w^T x_i + b)
\]  
\[(8)\]

The functional margin of a corpus of decision boundary is twice the functional margin of any of the samples in the corpus with minimal functional margin. The polynomial kernel with degree \(d\) of the SVM was implemented in this study and it is defined in Equation (9).\(^ {54}\)

\[
K(x_i, x_j) = (x_i^T x_j + 1)^d
\]  
\[(9)\]

2.2 Experimental procedure

From the various literatures studied, no suitable dataset for this task was found, which necessitated building our own dataset. Data collection was done at one of the laboratory studios in the Department of Systems Engineering, University of Lagos, Akoka, Lagos Nigeria. The laughter recordings were done in sessions spanning over 5 months (February to June 2018) and have been constantly upgraded. Unlike speech recordings, laughter bouts from individuals are not easily acquired in large quantity for research purpose. It also has a high degree of variability in types and forms. Participants in the recordings include undergraduate, post-graduate students, as well as random University staff across six faculties (Engineering, Sciences, Social Sciences, Arts, Environmental Sciences and Education). Age 17 and above were considered. Age 17 was chosen as the lowest age because research shows that the voice of a teenager becomes fully stable from that age having attained puberty.\(^ {55}\) More than 10 laughter samples of average of 3 seconds of 82 males and 41 females (a total of 123 participants), with a total of 1530 laughter samples were used in this study. The dataset is available at http://laughter-db.herokuapp.com/.

Figure 1 showed the procedure for the data acquisition, pre-processing, feature extraction, training, testing, and validation of the GMM and SVM models. The acoustic features used in literature\(^ {56,57}\) to discriminate laughter from speech were implemented in this study. These features include mean of intensity, pitch (fundamental frequency [F0]) and formants frequencies (F1, F2, F3, F4, and F5), with a view to distinguish male laughter from female laughter.

Twelve MFCC coefficients and energy were extracted and a dynamic average was carried out to generate the Dynamic-Average of the Mel frequency Cepstral Coefficient (DA-MFCC) in order to improve on the performance of the system. Both the raw MFCC and the DA-MFCC were then used for training and classification. This study was implemented in python programming language via the Scientific Python Development Environment (SPYDER) IDE of the Anaconda software.\(^ {58}\) To evaluate the proposed method, the Voxforge speech dataset a well-known dataset for speaker identification tasks\(^ {59}\) was used for gender identification. Ten speech samples from 32 females and 58 males were used for training while five samples from each individual were used for validation.

![Block diagram describing the methodology](image)

**FIGURE 1** Block diagram describing the methodology
3 | RESULTS AND DISCUSSION

The results of the various acoustic analysis carried out were reported in Table 1. This revealed the differences between spontaneous laughter and speech while Table 2 showed the differences between male and female laughter.

The MFCC and the DA-MFCC introduced in this study were extracted for laughter and Voxforge datasets, and thereafter used for the training, testing, and validation using the GMM and SVM classifiers. The results obtained were presented in Tables 3 to 6. Figures 2 and 3 revealed the differences in the training time for the MFCC and the DA-MFCC with the GMM and SVM respectively. The Precision-Recall and the confusion matrix for the Voxforge dataset with the SVM classifier were shown in Tables 7 to 9, while Tables 10 to 12 revealed the Precision-Recall and the confusion matrix for the laughter dataset. Table 13 showed the Precision-Recall with the GMM classifier.

Table 1 revealed that pitch (F0) and all the formant frequencies were higher in laughter than in speech. This is because there is an increase in airflow through the vocal tract for laughter, which results in faster vibration of the vocal folds. On the contrary, the intensity was higher in speech than in spontaneous laughter. This may be due to the fact that the energy of each bout of laughter reduces as spontaneous laughter progresses. The energy in speech may also increase as the word is being produced especially for emphasized words. This finding is consistent with the study reported by Truong and Leeuwen.45

In Table 2, the differences between male and female laughter were reported. Pitch (F0), second, and third formant frequencies (F2 and F3) of laughter were higher in females than in males, because females have shorter vocal folds (between 12.5 and 17.5 mm) than males (between 17 and 25 mm). According to the National Centre for Voice and Speech,60 the fundamental frequency is inversely proportional to the vocal fold length. Shorter folds produce higher frequencies, while longer folds produce lower frequencies. The findings of this study were consistent with observations of Bachorowski and Owren.57 It was observed also that the first, fourth, and fifth formant frequencies (F1, F4, and F5) of the males were higher

| Features | Laughter | Speech |
|----------|----------|--------|
| F0_Mean  | 336.69   | 191.86 |
| F1_Mean  | 909.76   | 706.93 |
| F2_Mean  | 2021.95  | 1947.26|
| F3_Mean  | 3016.77  | 2963.31|
| F4_Mean  | 4056.27  | 4045.08|
| F5_Mean  | 4862.17  | 4854.07|
| Intensity Mean | 50.70 | 54.04 |

| Features | Laughter male | Laughter female |
|----------|---------------|-----------------|
| F0_Mean  | 305.29        | 367.65          |
| F1_Mean  | 929.50        | 890.30          |
| F2_Mean  | 2002.82       | 2040.81         |
| F3_Mean  | 2982.98       | 3050.08         |
| F4_Mean  | 4076.93       | 4035.92         |
| F5_Mean  | 4913.99       | 4817.04         |
| Intensity Mean | 52.37 | 49.06 |

| Gender | Laughter testing accuracy (%) | Time taken to train laughter (s) | Laughter validation accuracy (%) | Voxforge dataset testing accuracy (%) | Time taken to train Voxforge (s) | Voxforge dataset validation accuracy (%) |
|--------|-------------------------------|----------------------------------|----------------------------------|--------------------------------------|----------------------------------|-------------------------------------------|
| Female | 80.28                         | 73.00                            | 80.78                            | 93.09                                | 76.70                            | 93.79                                     |
| Male   | 82.24                         | 76.00                            | 84.41                            | 83.52                                | 78.40                            | 80.10                                     |
| Overall accuracy | 81.26 | 74.50 | 82.60 | 88.31 | 77.55 | 86.95 |
TABLE 4  Training, testing and validation with GMM using DA-MFCC

| Gender  | Laughter testing accuracy (%) | Time taken to train laughter (s) | Laughter validation accuracy (%) | Voxforge dataset testing accuracy (%) | Time taken to train Voxforge (s) | Voxforge dataset validation accuracy (%) |
|---------|-------------------------------|----------------------------------|----------------------------------|----------------------------------------|----------------------------------|------------------------------------------|
| Female  | 82.26                         | 60.60                            | 84.00                            | 95.02                                  | 63.60                            | 95.12                                    |
| Male    | 91.06                         | 66.60                            | 91.30                            | 85.42                                  | 65.40                            | 80.20                                    |
| Overall | 86.66                         | 63.60                            | 87.65                            | 90.22                                  | 64.50                            | 87.66                                    |

TABLE 5  Training, testing, and validation with SVM using MFCC

| Gender  | Laughter testing accuracy (%) | Time taken to train laughter (s) | Laughter validation accuracy (%) | Voxforge testing accuracy (%) | Time taken to train Voxforge (s) | Voxforge validation accuracy (%) |
|---------|-------------------------------|----------------------------------|----------------------------------|------------------------------|----------------------------------|---------------------------------|
| Female  | 71.17                         | 1124.40                          | 81.00                            | 83.31                        | 1500.00                          | 90.09                           |
| Male    | 71.17                         | 1357.80                          | 86.88                            | 83.31                        | 1708.20                          | 94.79                           |
| Overall | 71.17                         | 1241.10                          | 83.94                            | 83.31                        | 1604.10                          | 92.44                           |

TABLE 6  Training, testing, and validation with SVM using DA-MFCC

| Gender  | Laughter testing accuracy (%) | Time taken to train laughter (s) | Laughter validation accuracy (%) | Voxforge testing accuracy (%) | Time taken to train Voxforge (s) | Voxforge validation accuracy (%) |
|---------|-------------------------------|----------------------------------|----------------------------------|------------------------------|----------------------------------|---------------------------------|
| Female  | 75.49                         | 810.00                           | 82.19                            | 88.15                        | 1353.40                          | 92.78                           |
| Male    | 75.49                         | 1087.20                          | 91.63                            | 88.15                        | 1042.20                          | 96.00                           |
| Overall | 75.49                         | 948.60                           | 86.91                            | 88.15                        | 1197.8                           | 94.39                           |

FIGURE 2  Time graph for training MFCC and DA-MFCC using the GMM classifier

than the females. The F1 was different from the result reported by Bachorowski and Owren, in that it was higher in females than in males while F4 and F5 corroborated the report by Bachorowski and Smoski. F4 indicated that 60% of our laughter were made up of voiced closed mouth.

Pitch (F0) and formant frequencies (F1 to F4) were higher in females’ speech than in males while F5 and intensity were higher in males’ speech than in females. This may be due to the fact that males have thicker vocal cords than females. This result was consistent with the studies by Bachorowski and Smoski.

Tables 3 and 4 revealed results for GMM while Tables 5 and 6 showed the results for the SVM classifier with MFCC and DA-MFCC respectively. The testing accuracy of the laughter dataset increased from 81.26% to 86.66% and from 71.17% to 75.49% for GMM and SVM classifiers respectively. Similarly, the validation accuracy also increased from 82.60% to 87.65% and from 83.94% to 86.91% for the GMM and SVM classifiers respectively when the DA-MFCC was implemented.
Furthermore, the overall training time reduced from 74.50 to 61.80 seconds and from 1241.10 to 948.60 seconds for the GMM and SVM respectively when the DA-MFCC was used. From the literature, a total of 39 coefficients consisting of the first 12 MFCC features and energy, as well as the first and second derivatives have been found to be significant in laughter\textsuperscript{61} and speech\textsuperscript{53} analyses. The shared similarities in syllables, frequencies, and root mean square amplitude for both laughter and speech have been reported in Gosztolya et al\textsuperscript{62} and Bickley and Hunnicutt.\textsuperscript{63} This appears reasonable since both are produced by the same vocal tract. The extensive analysis in this study using the dynamic-average MFCC approach (a modification of MFCC), reveals that of the 39 significant coefficients, the minimum range required to optimally capture the characteristic signatures in laughter are the first 11 coefficients. This interestingly is a little lower than the minimum range of 12 to 20 coefficients of MFCC required in voice-based recognition systems as reported in Akhtar.\textsuperscript{64}

For Voxforge dataset, the testing accuracy increased from 88.31% to 90.22% and from 83.31% to 88.15% for the GMM and SVM classifiers respectively. The validation accuracy also increased from 86.95% to 87.66% and from 92.44% to 94.39%
for the GMM and SVM classifiers respectively with the implementation of the DA-MFCC. Our method performed better than the accuracy reported by Gyanendra \(^9\) where 73\% accuracy was reported for 80 speakers.

Figures 2 and 3 respectively revealed the time graph analysis for the reduction in training time for the GMM and SVM classifiers when the DA-MFCC was used. There were 17.05\% and 23.59\% reduction in training time with the GMM and SVM respectively on our laughter dataset. Similarly, 16.83\% and 25.33\% reduction in training time were observed with the GMM and SVM respectively on the Voxforge dataset. This revealed that the DA-MFCC, uses less time for training the classifiers. The Precision-Recall and the confusion matrix for the Voxforge and laughter datasets were reported in Tables 7 to 13. There were high Precision, Recall, and F1-score for the two datasets. A total of 86.90\% and 74.32\% females were correctly classified for the Voxforge and laughter dataset respectively, while 13.10\% and 25.68\% were incorrectly classified for the Voxforge and laughter dataset respectively. In males, a total of 10.81\% and 23.23\% were incorrectly classified for the Voxforge and laughter dataset respectively, while 89.19\% and 76.77\% were correctly classified for the Voxforge and laughter datasets respectively.

### 4. CONCLUSION

The use of a vocalized form of human expression apart from speech, which has some features that can be harnessed in the task of GR has been explored in this study. Laughter and its paralinguistic features were introduced as a novel biometric trait for gender classification. Laughter cannot be controlled once it is aroused, it therefore, may be more beneficial when established for forensic use. Speech/voice can be mimicked perfectly but laughter can hardly be mimicked. In this study, spontaneous laughter captured from a total of 123 volunteers (82 males and 41 females) were analyzed and the Dynamic-Average MFCC features were extracted to train two classifiers namely, GMM and SVMs. Overall accuracies of 87.65\% and 83.94\% were obtained with GMM and SVM respectively, indicating the feasibility of using features extracted from laughter for identifying gender. For the SVM, model evaluation metrics of 0.76 precision revealed a high percentage of results were relevant and a recall of 0.76 showed a high percentage of total relevant results were correctly classified by the algorithm. The GMM algorithm also had a precision value of 0.83 and recall of 0.91.

There were 17.05\% and 23.59\% reduction in training time with the GMM and SVM respectively using this laughter dataset. Similarly, 16.83\% and 25.33\% reduction in training time were recorded with the GMM and SVM respectively when DA-MFCC was used on the Voxforge dataset. This study has demonstrated that human laughter is a viable biometric trait. The study also demonstrated that performance of the classifiers when using vocally generated sounds can be improved by applying the dynamic average to extend the traditional MFCC features. Future studies will seek to improve the model through the acquisition of more volunteers for the dataset, and utilize recent developments in deep learning algorithm.
PEER REVIEW INFORMATION

Engineering Reports thanks the anonymous reviewers for their contribution to the peer review of this work.

CONFLICT OF INTERESTS

The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS

Comfort Folorunso: Data curation; formal analysis; investigation; methodology; writing-original draft; writing-review and editing. Oluwatoyin Popoola: Conceptualization; project administration; supervision; writing-review and editing. Olumuyiwa Asaolu: Conceptualization; project administration; supervision; writing-review and editing.

ORCID

Comfort O. Folorunso https://orcid.org/0000-0003-1067-284X
Oluwatoyin P. Popoola https://orcid.org/0000-0001-5438-2807

REFERENCES

1. Folorunso CO, Asaolu OS, Popoola OP. Laughter signature: a novel biometric trait for person identification. Int J Biometrics. 2020;12(3):283-300.
2. Hamidi F, Scheuerman MK, Branham SM. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. Paper presented at: Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems; Apr 19, 2018:1-13.
3. Heredia-Jimenez J, Orantes-Gonzalez E. Gender differences in patients with fibromyalgia: a gait analysis. Clin Rheumatol. 2019;38(2):513-522.
4. Dehghan A, Ortiz EG, Shu G, Dager MSZ. Deep age, gender and emotion recognition using convolutional neural network; Feb 14, 2017. arXiv preprint arXiv:1702.04280.
5. Lemley J, Abdul-Wahid S, Banik D, Andonie R. Comparison of recent machine learning techniques for gender recognition from facial images. MAICS 2016. Dayton, OH, USA: Proceedings of the 27th Modern Artificial Intelligence and Cognitive Science Conference; 2016:97-102.
6. Mansanet J, Albiol A, Paredes R. Local deep neural networks for gender recognition. Pattern Recogn Lett. 2016;70:80-86.
7. Ranjan R, Patel VM, Chellappa R. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans Pattern Anal Mach Intell. 2017;41(1):121-135.
8. Kala N, Bhatia T, Aggarwal N. Person identification and characterization from gait using smartphone. Paper presented at: 2019 11th International Conference on Communication Systems & Networks (COMSNETS); Jan 7, 2019: IEEE:492-495.
9. Bianconi F, Smeraldi F, Abdollahyan M, Xiao F. On the use of skin texture features for gender recognition: An experimental evaluation. Paper presented at: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA); Dec 12, 2016: IEEE:1-6.
10. Tsimperidis I, Arampatzis A, Karakos A. Keystroke dynamics features for gender recognition. Digit Investig. 2018;24:4-10.
11. Kanani I, Shah H, Mankad SH. On the performance of Cepstral features for voice-based gender recognition. Information and Communication Technology for Intelligent Systems. Singapore, Singapore: Springer; 2019:327-333.
12. Kumar P, Baheeti P, Jha RK, Sarmah P, Sathish K. Voice gender detection using gaussian mixture model. J Netw Commun Emerg Technol. 2018;8(4):132-136.
13. Zvarevase K, Olugbara OO. Gender voice recognition using random forest recursive feature elimination with gradient boosting machines. Paper presented at: 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD); Aug 6, 2018: IEEE:1-6.
14. Aggarwal G, Vig R. Acoustic methodologies for classifying gender and emotions using machine learning algorithms. Paper presented at: 2019 Amity International Conference on Artificial Intelligence (AICAI); Feb 4, 2019: IEEE:672-677.
15. Kuppusamy K, Eswaran C. An advance on gender classification by information preserving features. Paper presented at: Proceedings of the 2018 International Conference on Electronics and Electrical Engineering Technology; Sep 19, 2018:227-231.
16. Pahwa A, Aggarwal G. Speech feature extraction for gender recognition. Int J Image Graph Signal Process. 2016;8(9):17-25.
17. Safavi S, Russell M, Jancović P. Automatic speaker, age-group and gender identification from children’s speech. Comput Speech Lang. 2018;50:141-156.
18. Shahin I, Nassif AB, Hamsa S. Emotion recognition using hybrid Gaussian mixture model and deep neural network. IEEE Access. 2019;7:26777-26787.
19. Tuncer T, Dogan S. Novel dynamic centre based binary and ternary pattern network using M4 pooling for real world voice recognition. Appl Acoust. 2019;156;176-185.
20. Tuncer T, Dogan S. A novel octopus-based Parkinson’s disease and gender recognition method using vowels. Appl Acoust. 2019;155;75-83.
21. Yaman O, Ertam F, Tuncer T. Automated Parkinson’s disease recognition based on statistical pooling method using acoustic features. Med Hypotheses. 2020;135:109483.
54. Chamasemani FF, Singh YP. Multi-class support vector machine (SVM) classifiers—an application in hypothyroid detection and classification. Paper presented at: 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications; Sep 27, 2011: IEEE:351-356.

55. Cleveland_Clinic. Voice changes: What can they tell you as you age? Online Health Essential website; 2017. https://health.clevelandclinic.org/voice-changes-what-can-they-tell-you-as-you-age/. Accessed May 5, 2020.

56. Bachorowski JA, Smoski MJ, Owren MJ. The acoustic features of human laughter. J Acoust Soc Am. 2001;110(3):1581-1597.

57. Bachorowski JA, Owren MJ. Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. Psychol Sci. 2001;12(3):252-257.

58. Raybaut, P. Spyder – Documentation Release 3. 2017. https://www.academia.edu/34872214/Spyder_Documentation_Release_3

59. Verma GK. Multi-feature fusion for closed set text independent speaker identification. International Conference on Information Intelligence, Systems, Technology and Management. Berlin, Germany: Springer; 2011:170-179.

60. National Centre for Voice and Speech. Voice classification and life-span changes; 2012. http://www.ncvs.org/ncvs/tutorials/voiceprod/equation/chapter7/index.html. Accessed June 1, 2020.

61. Salamin H, Polychroniou A, Vinciarelli A. Automatic detection of laughter and fillers in spontaneous mobile phone conversations. Paper presented at: 2013 IEEE International Conference on Systems, Man, and Cybernetics; 2013:IEEE:4282-4287.

62. Gosztolya G, Beke A, Neuberger T, Tóth L. Laughter classification using deep rectifier neural networks with a minimal feature subset. Arch Acoust. 2016;41:669-682.

63. Bickley CA, Hunnicutt S. Acoustic analysis of laughter. Second International Conference on Spoken Language Processing. Alberta, Canada: Proceedings of the International Conference on Spoken Language Processing; 1992.

64. Akhtar, M. “Text Independent Biometric Authentication System Based on Voice Recognition”, 2017. https://is.muni.cz/th/r7joz/MS-Thesis-459199.pdf.

How to cite this article: Folorunso CO, Popoola OP, Asaolu OS. Laughter signature, a new approach to gender recognition. Engineering Reports. 2020;2:e12267. https://doi.org/10.1002/eng2.12267