Animal Recognition using Cross-Correlation

Monish M. Shah[1], Mrs. Sangeeta Kulkarni[2]
[1, 2] K.J. Somaiya College of Engineering.Mumbai

Abstract- *Speaker recognition and identification is considered as the major problem since ages. Many researchers have studied the human's speech and analyzed to recognize and identify the speaker. A very less work has been done in the field of animal's sound voice pattern identification. There is an acute need for speech processing in order to recognize the animals of interest as we are penetrating into the wildlife civilization area day by day. The study focuses on the development of a speaker recognition system comprising a cross correlation technique. Over 100voice samples of various animals and birds have been considered for the demonstration purpose, noise is added to it and then tried to recognize the animal from the database. By addition of various types of noise we have tried to make the program robust to various types of natural noises that we usually encounter in natural environmental cases.*

Keywords- spatial-temporal, cross-correlation, FFT.

I. INTRODUCTION

Speaker recognition is the process of automatically recognizing who is speaking by using the speaker-specific information included in speech waves to verify identities being claimed by people accessing systems; that is, it enables access control of various services by voice.

The voice is considered both a physiological and a behavioral biometric factor:

- The physiological component of speaker recognition is the physical shape of the subject's voice tract;
- The behavioral component is the physical movement of jaws, tongue and larynx.

There are many interactions between the signal, the auditory system, and the central nervous system.In this paper the noisy sound is used as an input to the system and then efforts are made to identify the voice of the animal from the database. To increase the reliability we have tried to add various types of noises such as AWGN, random noise etc. which resembles to the noise added in the environmental conditions and then further applied the methods to identify the animal.

In a user interface it may be desirable to present multiple digitized speech recordings simultaneously, providing lateral capabilities while avoiding the time block intrinsic in speech communication because of the serial nature of audio [1, 2].

II. LITERATURE SURVEY

I] Principles of Speaker Recognition

Speaker recognition can be classified into speaker identification and speaker verification. Speaker identification is the process of determining from which of the registered speakers a given sound comes. Speaker verification is the process of accepting or rejecting the identity requested by a speaker. Most of the applications in which voice is used to confirm the identity of a speaker are classified as speaker verification.

In the speaker identification task, a speech sound from an unknown speaker is analyzed and compared with speech models of known speakers. The unknown speaker is identified as the speaker whose model best matches the input sound. In speaker verification, an identity is claimed by an unknown speaker, and a sound of this unknown speaker is compared with a model for the speaker whose identity is being claimed. If the match is good enough, that is, above a threshold, the identity claim is accepted.

The fundamental difference between identification and verification is the number of decision substitutes. In identification, the number of decision substitutes is equal to the size of the population, whereas in verification there are only two choices, acceptance or rejection, regardless of the population size. Therefore, speaker identification performance decreases as the size of the population increases, whereas speaker verification performance approaches a constant independent of the size of the population, unless the distribution of physical characteristics of speakers is extremely biased.[3, 4, 5]

II] Text-Dependent and Text-Independent Methods

There exist two types of speaker recognition methods viz., text-dependent (fixed passwords) and text-independent (no specified passwords) methods. The previous require the speaker to provide sounds of key words or sentences, the same text being used for both training and recognition, whereas the

second do not depend on a specific text being spoken. The text-dependent methods are usually based on pattern/model-sequence-matching techniques in which the time axes of an input speech sample and reference templates or reference models of the registered speakers are ranged, and the similarities between them are accrued from the beginning to the end of the sound. Text dependent recognition has better performance for subjects that cooperate. But text independent voice recognition is more flexible that it can be used for non-cooperating individuals.[3, 4, 5] In this paper text dependent speaker recognition task is considered.

III. CROSS-CORRELATION

The cross correlation function measures the dependence of the values of one signal on another signal. For two WSS (Wide Sense Stationary) processes that are continuous in nature, $x(t)$ and $y(t)$ it is described by:

$$R_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}x(t)\,y(t+\tau)\,dt \quad Or$$

$$R_{yx}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}y(t)\,x(t+\tau)\,dt \qquad (i)$$

Where T is the observation time.

For sampled signals, it is defined as:

$$R_{yx}(m)=\frac{1}{N}\sum_{n=1}^{N-m+1}y(n)\,x(n-m+1) \quad Or$$

$$R_{yx}(m)=\frac{1}{N}\sum_{n=1}^{N-m+1}y(n)\,x(n-m+1) \qquad (ii)$$

m=1,2,3,..,N+1
Where N is the record length (i.e. number of samples).

IV. PROPOSED ALGORITHM

In this section, we describe the algorithm based on our proposed method for the consideration of cross correlation for the task of identification of animal from the database. Following Block diagram explain the process of animal recognition followed in this paper:

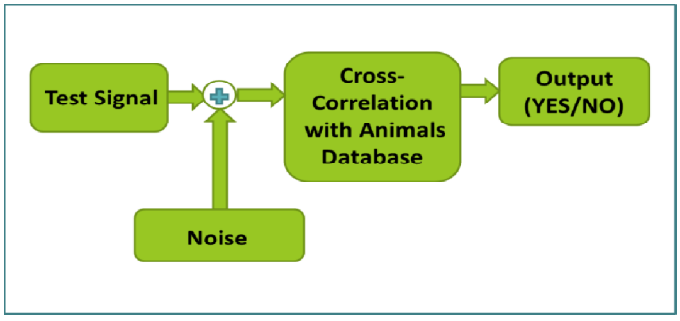

Figure 1. Block Diagram of Animal Recognition using Cross-correlation.

1. The features of animal's voices such as Amplitudes, Pitches, Formantsand Amplitude etc. of 10 Animals sample have been studied using PRAAT software.
2. The same 10 animals have been tested using Cross-correlation to identify the unknown animal from the dynamically generated database. The results along with the efficiency have been mention in detail later in this paper.
3. Addition of various types of noise such as Sawtooth Noise Function, Random Noise Function, Ceil Function Noise, Floor Function Noise and AWGN etc. to the input signal is done and then tried to identify the animal from the database.
4. Further we have extended the applicability of the research by increasing the database ten times larger than the previous i.e. a database of 100 animals have been created and the efficiency and results have been mentioned later in this paper.

V. RESULTS

Following is the image of the sparrow's sound. However all 100 sounds of animals and birds have been observed and has helped to draw the conclusion for the signals of the animals and birds that are not being detected by our methods. The figure given below shows spatial-temporal representation of speech signal. Time is represented on X-axis and normalized frequency on Y-axis. Second subplot shows Intensity of the speech signal.

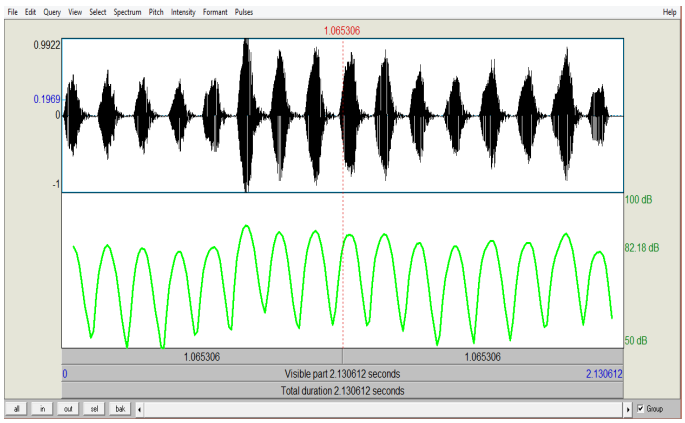

Figure 2. Spatio-temporal plot of Sparrow.

The following table shows the features that have been extracted from the speech signals using PRAAT. It shows Time duration, Sampling Frequency, Total Energy and Number of Samples of 10 Animals. However we have extracted the similar feature for the rest of the 90 animals too.

Table 1.Time duration, Sampling Frequency, Total Energy and Number of Samples of 10 Animals

Animal Name	Duration of Sample (sec)	Sampling Frequency (Hz)	Total Energy (Pascal² sec)	Number of Samples
American Robin	3.43	44100	0.06186	151359
Bear	4.53	11025	0.30627	49981
Sparrow	2.13	11025	0.09341	23490
Bison	3.84	11025	0.29097	42420
Blue jay	7.92	44100	0.52282	349432
Cheetah	6.48	22050	0.02942	142976
Sheep	0.99	22050	0.02227	21856
Geese	5.93	11025	0.17923	65436
Monkey	1.70	44100	0.03036	118994
Tiger	4.53	11025	0.17524	49946

For the process of identification we have created a database of 100 animals and then detected whether the particular signal belongs to that database or not. In this case we have added noise to the input signal. This noise has been added using AWGN function. The audio effect of the test signal then resembles the natural noise added in the environment and thus creating a perception of another sort of the practical cases.SNR used in this particular case is –10dB and –4dB. However we have tested the efficiency of animals testing results by varying the values of SNR from +25dB to –25dB and have plotted the graph for the same.

The table has given below shows the results for only two different values of SNR and that too for just a database of 100 animals; however we tested it for 100 different animals and birds whose efficiency is summarized later.

Table 2.Animal Identification from 10 Animals Database(AWGN Function: SNR = –10dB, –4dB)

Sr. No.	Name of the Animal	Random Function Noise	Ceil Function Noise	Floor Function Noise	SNR = –10dB	SNR = –4dB
		Match?	Match?	Match?	Match?	Match?
1	Sparrow	Yes	Yes	Yes	Yes	Yes
2	Tiger	Yes	Yes	Yes	Yes	Yes
3	Bison	Yes	Yes	Yes	Yes	Yes
4	Bear	Yes	Yes	Yes	Yes	Yes
5	American Robin	Yes	Yes	Yes	Yes	Yes
6	Cheetah	Yes	Yes	No	No	Yes
7	Geese	Yes	Yes	Yes	Yes	Yes
8	Monkey	Yes	Yes	No	No	Yes
9	Sheep	Yes	Yes	Yes	Yes	Yes
10	Bluejay	Yes	Yes	Yes	Yes	Yes

The graph shown below indicates that how the efficiency drops from 100 percent to zero by varying the values of SNR, which gives us an idea of till what of noise can the particular animal be detected.

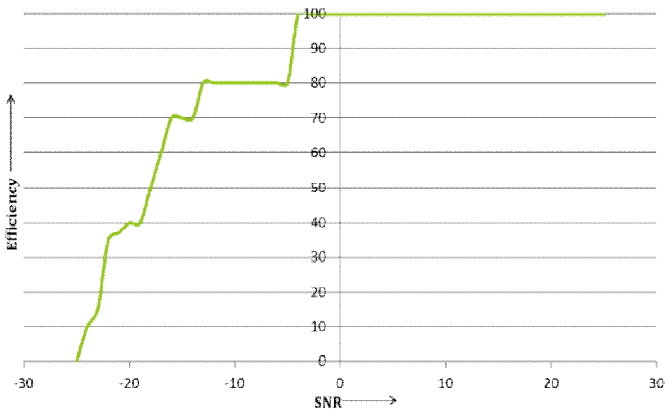

Figure 3. Variation in Efficiency of Animal recognition with changing values of SNR

VI. CONCLUSIONS

Following are the conclusions when the verification process is done for more than 100 animals:

1. Animal Identification task – efficiency is 78 percent and average time for simulation is 8 seconds.
2. Using Saw tooth Function Noise – efficiency is 82 percent average time for simulation is still 8 seconds.
3. Using Random Function Noise efficiency is 72 percent average time for simulation is again 8 seconds.

4. Using AWGN of SNR = –4dB efficiency is 61 percent average time for simulation is still 8 seconds.
5. It has been observed that some of the animals or birds are not identified when the input given to the system is without adding the noise since their frequency variations are very low and the modulus value of mean normalized signals is less than 1.

REFERENCES

[1] A. S. Bregman. "Auditory Scene Analysis: The Perceptual Organization of Sound". *MIT Press, 1990.*

[2] B. Arons. "Hyperspeech: Navigating in speech-only hypermedia." *Hypertext '91, pages 133-146. ACM, 1991.*

[3] Furui, S. (1991) "Speaker-Independent and Speaker-Adaptive Recognition Techniques," in Furui, S. and Sondhi, M. M. (Eds.) Advances in Speech Signal Processing, New York: Marcel Dekker, pp. 597-622.

[4] Furui, S. (1997) "Recent Advances in Speaker Recognition", Proc. First Int. Conf. Audio- and Video-based Biometric Person Authentication, Crans-Montana, Switzerland, pp. 237-252.

[5] Furui, S. (2000) Digital Speech Processing, Synthesis, and Recognition, 2nd Edition, New York: Marcel Dekker.

[6] For Birds and Animals Sounds Database following websites have been referred:
www.animal-sounds.org/
https://www.freesoundeffects.com/free-sounds/animals-10013/
www.soundbible.com/tags-animal.html/
www.audiomicro.com/free-sound-effects/free-animal-sound-effects
www.xeno-canto.org/
www.findsounds.com/isapi/
www.soundsnap.com/tags/thailand